SOLUTION-FREE SETS FOR SUMS OF BINARY FORMS

SEAN PRENDIVILLE

Abstract. In this paper we obtain quantitative estimates for the asymptotic density of subsets of the integer lattice $\mathbb{Z}^2$ which contain only trivial solutions to an additive equation involving binary forms. In the process we develop an analogue of Vinogradov's mean value theorem applicable to binary forms.

1. Introduction

Certain systems of linear equations have the property that, should a set of integers fail to deliver non-trivial solutions to the system, then the set has zero density. The problem of obtaining quantitative asymptotic estimates for the density of such sets, first addressed successfully by Roth [14, 15], is one which has seen remarkable advances over the last decade; spectacularly in the work of Gowers [6] on long arithmetic progressions, Bourgain [3, 4] on progressions of length three and Green and Tao [7] on progressions of length four. Recently, M. L. Smith [19] has obtained density estimates for sets of integers which do not contain solutions to a class of homogeneous equations involving kth powers. This was the first general result on inherently non-linear systems. In this paper we not only generalise Smith's result from an equation involving kth powers to one involving binary forms, but we also extract density estimates for subsets of the two-dimensional integer lattice. Our approach uses the density increment method of Roth and Gowers, together with the circle method. A notable feature in our application of the circle method is a novel analogue of Vinogradov's mean value theorem, applicable to systems of equations involving binary forms. Our approach to this mean value theorem makes intrinsic use of the structure of the shift-invariant system associated with our equation, and thereby improves on those estimates which can be deduced from the much more general work of Parsell [13] on multi-dimensional versions of Vinogradov's mean value theorem.

In order to describe our conclusions, we first introduce some notation. When $\Phi \in \mathbb{Z}[x, y]$ is a binary form we write $\Phi^{u,v}$ for the derivative $\frac{\partial^{u+v}}{\partial x^u \partial y^v}\Phi(x, y)$.

Definition 1.1. Let us say the tuple $\mathbf{c} = (c_1, \dots, c_s)$ of non-zero integers is a *non-singular choice of coefficients* for Φ if there exist binary forms $\Phi_1, \dots, \Phi_N$ satisfying

$$\begin{aligned}\{\Phi_1, \dots, \Phi_N\} &\subset \{\Phi^{u,v} : 0 \le u + v < k\}\\ &\subset \ \mathrm{span}\{\Phi_1, \dots, \Phi_N\},\end{aligned} \tag{1.1}$$

such that the auxiliary system of equations

$$c_1\Phi_i(\mathbf{x}_1) + \cdots + c_s\Phi_i(\mathbf{x}_s) = 0 \quad (1 \le i \le N), \tag{1.2}$$

2000 *Mathematics Subject Classification.* Primary 11P55; Secondary 11D45.

During the completion of this work, the author was supported by an EPSRC doctoral training grant through the University of Bristol.

has non-singular[1] real and p-adic solutions for every prime p.

Definition 1.2. We call a $2s$-tuple $(\mathbf{x}_1, \dots, \mathbf{x}_s)$ *diagonal* if there exists an affine line $L = \mathbf{a} + \mathbb{R} \cdot \mathbf{b}$ such that $\mathbf{x}_i \in L$ for all i.

Writing $[X]$ for the set $\{1, 2, \dots, \lfloor X \rfloor\}$, the most accessible of our density results can now be stated.

Theorem 1.3. *Let $\Phi \in \mathbb{Z}[x, y]$ be a binary form of degree $k \geq 2$ and let $\mathbf{c} \in \mathbb{Z}^s$ be a non-singular choice of coefficients for Φ, with $c_1 + \cdots + c_s = 0$. Suppose that $s \geq \frac{3}{4}k^3 \log k(1 + o(1))$. Then any set $A \subset [X]^2$ containing only diagonal solutions to the equation*

$$c_1\Phi(\mathbf{x}_1) + \cdots + c_s\Phi(\mathbf{x}_s) = 0 \qquad (\mathbf{x}_i \in A), \tag{1.3}$$

satisfies the bound

$$|A| \ll X^2 \left(\log\log X\right)^{-1/(s-1)}, \tag{1.4}$$

where the implicit constant depends only on $\mathbf{c}$ and Φ.

Remark 1.4. For a more precise lower bound on the number of variables required than $s \geq \frac{3}{4}k^3 \log k(1 + o(1))$, see Theorem 5.1.

For comparison, recent work of Smith [18] establishes a version of the above result in which Φ is replaced by a kth power and the set A is a subset of the integers in the interval $[1, X]$. Indeed, our insistence that A contains only diagonal solutions to (1.3) precludes the deduction of Theorem 1.3 from Smith's result. We also note that Smith obtains an exponent of $\log\log N$ in (1.4) of the form -2^{-2^k}.

One can obtain a qualitative version of Theorem 1.3 by applying the multidimensional Szemerédi theorem of Furstenberg and Katznelson [5]. In this way, one can show that any (infinite) set $A \subset \mathbb{Z}^2$ containing only diagonal solutions to (1.3) must have zero upper Banach density. If one had a quantitative version of the multidimensional Szemerédi theorem providing bounds analogous to the one-dimensional bounds of Gowers [6], then one could use this result to obtain bounds of the form (1.4) in Theorem 1.3. However, the exponent of $\log\log N$ in these bounds would be intrinsically dependent on the choice of form Φ and coefficients $c_1, \dots, c_s$, whereas our result depends only on s. Moreover, no such two-dimensional bounds currently exist; the best bounds presently available are due to Shkredov [17] and are not general enough for our purposes.

To obtain Theorem 1.3, we bound the density of sets which contain only diagonal solutions to the larger system of equations

$$c_1\Phi^{u,v}(\mathbf{x}_1) + \cdots + c_s\Phi^{u,v}(\mathbf{x}_s) = 0 \qquad (0 \leq u + v < k). \tag{1.5}$$

Sets avoiding non-diagonal solutions to this larger system may have greater size than those avoiding non-diagonal solutions to (1.3). However, a key observation is that this larger system enjoys *translation-dilation invariance*, in that $(\mathbf{x}_1, \dots, \mathbf{x}_s)$ satisfies (1.5) if and only if $(\lambda\mathbf{x}_1 + \boldsymbol{\xi}, \dots, \lambda\mathbf{x}_s + \boldsymbol{\xi})$ satisfies (1.5), whenever $\lambda \neq 0$. This invariance allows us to adapt the density increment method of Roth and Gowers [14, 6].

In order to implement the density increment method it is necessary to have an asymptotic estimate for the number of solutions to (1.5) with variables restricted to the interval $[1, X]$. This we obtain through an application of the Hardy–Littlewood

[1] Here *non-singular* means the associated Jacobian has full-rank over the field in question.

method. In order to deal with the minor arcs, we utilise Vinogradov's method[2], which necessitates the estimation of the number $J_{s,\Phi}(X)$ of solutions $(\mathbf{x},\mathbf{y}) \in [X]^{4s}$ to the system of equations

$$\sum_{j=1}^{s} \Phi^{u,v}(\mathbf{x}_j) = \sum_{j=1}^{s} \Phi^{u,v}(\mathbf{y}_j) \quad (0 \leq u+v < k). \tag{1.6}$$

When Φ takes the form $a(bx+cy)^k$, such an estimate can be obtained from the standard Vinogradov mean value theorem, as found in [21, Chapter 5]. We must therefore treat the remaining case.

Definition 1.5. We say a binary form $\Phi \in \mathbb{Z}[x,y]$ of degree k is *degenerate* if it takes the form $(\alpha x + \beta y)^k$ for some $\alpha, \beta \in \mathbb{C}$. One can check that Φ is degenerate if and only if there exist $a,b,c \in \mathbb{Z}$ such that $\Phi = a(bx+cy)^k$.

Definition 1.6. We define the *differential dimension* of Φ to be the dimension N of the linear span of the set of non-constant derivatives

$$\{\Phi^{u,v} : 0 \leq u+v < k\}. \tag{1.7}$$

Given a maximal linearly independent subset $\{F_1, \ldots, F_N\}$ of (1.7), we define the *differential degree* of Φ to be the quantity

$$K = \sum_i \deg F_i. \tag{1.8}$$

Elementary linear algebra confirms that K is independent of our choice of F_i.

Our mean value theorem for non-degenerate binary forms is then the following.

Theorem 1.7. *Let $\Phi \in \mathbb{Z}[x,y]$ be a non-degenerate binary form of degree k, differential dimension N and differential degree K. Write $M = \lceil N/2 \rceil$, and define*

$$\Delta_s = K\left(1 - \tfrac{1}{k}\right)^{\lfloor s/M \rfloor}. \tag{1.9}$$

Then we have the bounds

$$X^{4s-K} \ll J_{s,\Phi}(X) \ll X^{4s-K+\Delta_s}, \tag{1.10}$$

where the implicit constants depend only on s and Φ.

We remark that when Φ is a degenerate binary form, then $K = k(k+1)/2$ and $N = k$. Hence our result is comparable to the standard Vinogradov mean value theorem, where one obtains

$$\Delta_s \leq \tfrac{1}{2}k^2\left(1 - \tfrac{1}{k}\right)^{\lfloor s/k \rfloor}.$$

Using very general work of Parsell [13], one can extract a bound on the exponent Δ_s in Theorem 1.7 of the form

$$\Delta_s \leq rk\, e^{2-2s/rk},$$

where $r = (k+2)(k+3)/2 - 1$. By way of comparison, an immediate consequence of Theorem 1.7 is the bound

$$\Delta_s \leq K e^{-\frac{1}{k}\lfloor 2s/(N+1) \rfloor},$$

and one certainly has $K < rk$ and $N < r$. Moreover, Parsell's general theorem is obtained through the somewhat formidable method of *repeated efficient differencing*. We are able to extract our result from the comparatively simple p-adic

[2] See Chapter 4 of [11] for a description of this method.

iterative method, originating with Linnik [10], and reaching a refined state in work of Karatsuba [9] and Stechkin [20].

An expert in the field might hope to apply the above result via Vinogradov's method to obtain superior bounds for exponential sums over binary forms, at least when k is large. However, as demonstrated in Wooley [23, §8], one can already attain such bounds using the standard Vinogradov mean valued theorem.

1.1. Notation.

Throughout the remainder of the paper we fix a non-degenerate binary form Φ of degree k, differential dimension N and differential degree K. We reserve the letter M for the quantity $\lceil N/2\rceil$. Let us also fix $\{F_1,\dots,F_N\}$, a maximal linearly independent subset of $\{\Phi^{u,v} : 0 \le u+v < k\}$. Let $\mathbf{F}$ denote the tuple $(F_1,\dots,F_N)$. Setting $k_i = \deg F_i$, we always assume that $k = k_1 \ge k_2 \ge \cdots \ge k_N = 1$. Using Taylor's formula, a convenient consequence of our ordering of the F_i is that for any $\boldsymbol{\xi} \in \mathbb{Z}^2$ there exists a lower unitriangular[3] matrix $\Xi_{\boldsymbol{\xi}} \in GL_N(\mathbb{Q})$ such that

$$\mathbf{F}(\mathbf{x}+\boldsymbol{\xi}) = \Xi_{\boldsymbol{\xi}} \cdot \mathbf{F}(\mathbf{x}) + \mathbf{F}(\boldsymbol{\xi}). \tag{1.11}$$

We call this property *translation-dilation invariance*, since it implies that for any $\boldsymbol{\xi} \in \mathbb{R}^2$ and $\lambda \ne 0$ we have the equivalence

$$\sum_{j=1}^{s}\Big(\mathbf{F}(\mathbf{x}_j) - \mathbf{F}(\mathbf{y}_j)\Big) = 0 \iff \sum_{j=1}^{s}\Big(\mathbf{F}(\lambda\mathbf{x}_j+\boldsymbol{\xi}) - \mathbf{F}(\lambda\mathbf{y}_j+\boldsymbol{\xi})\Big) = 0. \tag{1.12}$$

Given a real $X \ge 1$ write $[X]$ for $\{1,2,\dots,\lfloor X\rfloor\}$. We use $J_{s,\Phi}(X;\mathbf{m})$ to denote the number of $(\mathbf{x},\mathbf{y}) \in [X]^{4s}$ satisfying

$$\sum_{j=1}^{s}\Big(\mathbf{F}(\mathbf{x}_j) - \mathbf{F}(\mathbf{y}_j)\Big) = \mathbf{m}. \tag{1.13}$$

Notice that $J_{s,\Phi}(X;\mathbf{0})$ coincides with our definition of $J_{s,\Phi}(X)$.

We analyse both the equations (1.5) and (1.6) via the exponential sum

$$f(\boldsymbol{\alpha}) = f(\boldsymbol{\alpha};X) = \sum_{\mathbf{x}\in[X]^2} e(\boldsymbol{\alpha}\cdot\mathbf{F}(\mathbf{x})), \tag{1.14}$$

where $e(y) = e^{2\pi i y}$. By the orthogonality relations we have

$$J_{s,\Phi}(X;\mathbf{m}) = \oint |f(\boldsymbol{\alpha})|^{2s} e(\boldsymbol{\alpha}\cdot\mathbf{m})\mathrm{d}\boldsymbol{\alpha}, \tag{1.15}$$

where $\oint$ denotes the integral over the N-dimensional torus $\mathbb{T}^N = \mathbb{R}^N/\mathbb{Z}^N$.

Throughout, we assume that X is sufficiently large in terms of s, $\mathbf{c}$ and $\mathbf{F}$, so all implicit constants depend only on these parameters, unless otherwise indicated. We note that $\mathbf{F}$ depends ultimately only on Φ.

2. The Mean Value Theorem

Before working towards upper bounds for $J_{s,\Phi}(X)$, let us derive an elementary lower bound. By (1.15), for any $\mathbf{m}$ we have $J_{s,\Phi}(X;\mathbf{m}) \le J_{s,\Phi}(X)$. Notice that there are $O_{\mathbf{F},s}(X^K)$ values of $\mathbf{m}$ for which $J_{s,\Phi}(X;\mathbf{m})$ is non-zero. Summing over these values, we obtain

$$X^K J_{s,\Phi}(X) \gg X^{4s}. \tag{2.1}$$

The lower bound in (1.10) follows.

[3] A lower triangular matrix with all diagonal entries equal to one.

The remainder of this section is occupied with proving the upper bound in (1.10). We expect the majority of solutions to (1.6) to be non-singular (in a sense to be defined later), whilst the remaining set of singular solutions should be relatively sparse. To define the appropriate notion of singularity neccesitates the discussion of the Jacobian associated to (1.6).

Definition 2.1. Write $\mathrm{Jac}(\mathbf{x}_1,\dots,\mathbf{x}_M)$ for the $N \times 2M$ matrix

$$\Big(F_i^{1,0}(\mathbf{x}_j),\ F_i^{0,1}(\mathbf{x}_j)\Big)_{\substack{1\le i\le N\\ 1\le j\le M}}, \tag{2.2}$$

and let $\Delta(\mathbf{x}_1,\dots,\mathbf{x}_M)$ denote the determinant of the $N\times N$ matrix consisting of the first N columns of $\mathrm{Jac}(\mathbf{x}_1,\dots,\mathbf{x}_M)$.

In order to get our version of Linnik's p-adic iterative method to work, Δ cannot be identically zero. Notice that if Φ is degenerate, then Δ *is* identically zero. Our first lemma, Lemma 2.2, feeds into our second, Lemma 2.3, which establishes that Δ is non-zero when and only when Φ is non-degenerate. We keep Lemma 2.2 separate as it proves useful later.

Lemma 2.2. *Suppose there exists* $1 \le l < \deg\Phi$ *such that the linear span of the set* $\{\Phi^{u,v} : u+v=l\}$ *is one-dimensional. Then* Φ *is degenerate.*

Proof. For any binary form Φ of degree k and $0\le l\le k$ one can show by induction that

$$\Phi = \frac{(k-l)!}{k!}\sum_{r=0}^{l}\binom{l}{r}x^{l-r}y^r\Phi^{l-r,r}. \tag{2.3}$$

Let F_i be the only form from $F_1,\dots,F_N$ with degree $k-l$. Then for each $r=0,\dots,l$ there exists $\lambda_r\in\mathbb{Q}$ such that $\Phi^{l-r,r}=\lambda_rF_i$. We must have $\lambda_r\ne0$ for some r. Let us suppose that $r>0$, the case $r<l$ being similar. We have

$$F_i^{1,0}=\lambda_r^{-1}\Phi^{l-r+1,r}=\lambda_r^{-1}\lambda_{r-1}F_i^{0,1}.$$

Letting $\lambda=\lambda_r^{-1}\lambda_{r-1}$ and iterating one sees that for all $0\le s\le k_i$ we have $F_i^{k_i-s,s}=\lambda^{k_i-s}F_i^{0,k_i}$. Using this and (2.3), it follows that

$$\begin{aligned} F_i &= \frac{F_i^{0,k_i}}{k_i!}\sum_{r=0}^{k_i}\binom{k_i}{r}\lambda^{k_i-r}x^{k_i-r}y^r \\ &= (\alpha x+\beta y)^{k_i} \end{aligned}$$

for some real α and β, with $\beta\ne0$. Differentiating in y we see that

$$\Phi^{l-r,r+1}=\lambda_rF_i^{0,1}=\lambda_r\beta k_i(\alpha x+\beta y)^{k_i-1}.$$

Differentiating in x when $r<l$ we also see that

$$\Phi^{l-r,r+1}=\lambda_{r+1}F_i^{1,0}=\lambda_{r+1}\alpha k_i(\alpha x+\beta y)^{k_i-1}.$$

Thus for each $r=0,1,\dots,l$, one obtains

$$\lambda_r=(\alpha/\beta)\lambda_{r+1}=\dots=(\alpha/\beta)^{l-r}\lambda_l.$$

Inputting this into (2.3) we deduce that

$$\begin{aligned}\Phi &= \frac{l!}{k!}\sum_{r=0}^{l}\binom{l}{r}x^{l-r}y^{r}(\alpha/\beta)^{l-r}\lambda_l(\alpha x+\beta y)^{k-l}\\ &= \frac{l!\lambda_l}{k!\beta^l}(\alpha x+\beta y)^k .\end{aligned}$$

Therefore Φ is degenerate. □

Lemma 2.3. *If Φ is a non-degenerate binary form, then the determinant Δ is not the zero polynomial.*

Proof. For each $1 \le l \le k$, let $I(l)$ denote the set of indices i for which $k_i := \deg F_i = l$. For each $i \in I(l)$ there exists $\mathbf{c}_{li} \in \mathbb{Z}^{l+1}$ such that $F_i(x,y) = \mathbf{c}_{li}\cdot(x^l, x^{l-1}y, \dots, y^l)$. Let C_l denote the matrix whose rows comprise $\mathbf{c}_{li}$ $(i \in I(l))$. Since the F_i are linearly independent, C_l has full-rank. Hence there exists an invertible matrix B_l such that B_lC_l is a full-rank matrix in reduced row-echelon form. Define the rational homogeneous polynomials $G_1,\dots,G_N$ by

$$\begin{pmatrix} G_1\\ \vdots \\ G_N\end{pmatrix} = \begin{pmatrix} B_k & & & \\ & B_{k-1} & & \\ & & \ddots & \\ & & & B_1\end{pmatrix}\cdot\begin{pmatrix} F_1\\ \vdots \\ F_N\end{pmatrix}.$$

From our construction, we see that $\deg G_i = \deg F_i$ for all i. Furthermore, if d_i denotes the highest exponent of x occurring in $G_i(x,y)$, then for any $i,j \in I(l)$ with $i<j$ we have $d_i > d_j$. Write $\widetilde{\mathrm{Jac}}(\mathbf{x}_1,\dots,\mathbf{x}_M)$ for the $N \times 2M$ matrix

$$\Big(G_i^{1,0}(\mathbf{x}_j),\ G_i^{0,1}(\mathbf{x}_j)\Big)_{\substack{1\le i\le N\\ 1\le j\le M}}, \tag{2.4}$$

and let $\tilde{\Delta}(\mathbf{x}_1,\dots,\mathbf{x}_M)$ denote the determinant of its first N columns. By linearity of differentiation, we have

$$\widetilde{\mathrm{Jac}}(\mathbf{x}_1,\dots,\mathbf{x}_M) = \begin{pmatrix} B_k & & & \\ & B_{k-1} & & \\ & & \ddots & \\ & & & B_1\end{pmatrix}\cdot \mathrm{Jac}(\mathbf{x}_1,\dots,\mathbf{x}_M).$$

Since the matrix with the B_i along the diagonal is non-singular, it suffices to prove that $\tilde{\Delta}(\mathbf{x}_1,\dots,\mathbf{x}_M)$ is not the zero polynomial.

For r in the range $1 \le r \le \frac{N}{2}$, define $D_r(\mathbf{x}_1,\dots,\mathbf{x}_r)$ to be the determinant of the $2r\times 2r$ matrix occurring in the bottom-left corner of $\widetilde{\mathrm{Jac}}(\mathbf{x}_1,\dots,\mathbf{x}_M)$. We induct on r to show D_r is not the zero polynomial. When $N = 2M$ this completes the proof, since in this case $D_M = \tilde{\Delta}$. When $N+1 = 2M$ we expand along the Nth column of the matrix associated to $\tilde{\Delta}$ to obtain

$$\tilde{\Delta} = D_{M-1}(\mathbf{x}_1,\dots,\mathbf{x}_{M-1})G_1^{1,0}(\mathbf{x}_M) + \sum_{i=2}^{N} P_i(\mathbf{x}_1,\dots,\mathbf{x}_{M-1})G_i^{1,0}(\mathbf{x}_M), \tag{2.5}$$

for some polynomials $P_2,\dots,P_N$. Since F_1 is the only form of degree $k_1 = k$ and is non-degenerate, $G_1(x,y)$ does not take the form cy^{k_1}. It follows that $G_1^{1,0}(x,y)$ is a non-zero polynomial of degree $k-1$, a degree higher than that of any other $G_i^{1,0}(x,y)$. Since $D_{M-1}(\mathbf{x}_1,\dots,\mathbf{x}_{M-1})$ is also a non-zero polynomial, we can use

(2.5) to compare the exponents of the monomials in $\tilde{\Delta}$ which feature $\mathbf{x}_M$, and thereby deduce that $\tilde{\Delta}$ cannot be zero.

It remains to show that D_r is non-zero for each $1 \leq r \leq N/2$. We begin with a claim.

Claim. *For $2 \leq i \leq N$ the polynomial*

$$W_i(x,y) = \begin{vmatrix} G_{i-1}^{1,0} & G_{i-1}^{0,1} \\ G_i^{1,0} & G_i^{0,1} \end{vmatrix} = G_{i-1}^{1,0}\ G_i^{0,1} - G_{i-1}^{0,1}\ G_i^{1,0} \tag{2.6}$$

is non-zero, of degree $k_{i-1}+k_i-2$ and with highest exponent of x equal to $d_{i-1}+d_i-1$.

Recalling that d_i denotes the highest exponent of x occurring in $G_i(x,y)$, consider the polynomial

$$\begin{vmatrix} d_{i-1}x^{d_{i-1}-1}y^{k_{i-1}-d_{i-1}} & (k_{i-1}-d_{i-1})x^{d_{i-1}}y^{k_{i-1}-d_{i-1}-1} \\ d_i x^{d_i-1}y^{k_i-d_i} & (k_i-d_i)x^{d_i}y^{k_i-d_i-1} \end{vmatrix}$$
$$= \big(d_{i-1}(k_i-d_i) - d_i(k_{i-1}-d_{i-1})\big)x^{d_{i-1}+d_i-1}y^{k_{i-1}+k_i-d_{i-1}-d_i-1}.$$

If this is non-zero then, by our construction of the G_j, it has the same leading monomial and coefficient as W_i (when we order monomials according to the lexicographical[4]ordering on their exponents). To establish the claim it therefore remains to show that

$$\big(d_{i-1}(k_i-d_i) - d_i(k_{i-1}-d_{i-1})\big) \neq 0. \tag{2.7}$$

Suppose otherwise. Then

$$d_{i-1}k_i = d_i k_{i-1}. \tag{2.8}$$

There are two cases to consider. In the first case $k_i = k_{i-1}$, from which it follows that $d_{i-1} = d_i$. However, this contradicts our construction of the G_j. The only other possibility is that $k_{i-1} = k_i + 1$. In this case k_{i-1} and k_i are co-prime, so we must have $d_{i-1} = k_{i-1}$ and $d_i = k_i$. Our construction of the G_j therefore ensures that G_{i-1} is the only G_j of degree k_{i-1}, since it has the highest index of any G_j of degree k_{i-1}, but also has highest exponent of x equal to k_{i-1}. This forces Φ to be degenerate, by Lemma 2.2, a contradiction which establishes the claim.

Notice that $D_1(\mathbf{x}_1) = W_N(\mathbf{x}_1)$, so D_1 is a non-zero polynomial by the claim, giving us the basis case of our induction. Let us suppose that D_{r-1} is non-zero, with $r \leq N/2$. Inspection reveals that D_r is equal to

$$\sum_{N-2r<i<j\leq N} P_{ij}(\mathbf{x}_1,\dots,\mathbf{x}_{r-1})\big(G_i^{1,0}(\mathbf{x}_r)G_j^{0,1}(\mathbf{x}_r) - G_i^{0,1}(\mathbf{x}_r)G_j^{0,1}(\mathbf{x}_r)\big), \tag{2.9}$$

where the P_{ij} are polynomials with $P_{ij} = D_{r-1}$ when

$$\{i,j\} = \{N-2r+2, N-2r+1\}.$$

Let W_{ij} denote the polynomial $G_i^{1,0}G_j^{0,1} - G_i^{0,1}G_j^{0,1}$. The degree of the W_{ij} occurring in (2.9) is maximised only when $k_i = k_{N-2r+1}$ and $k_j = k_{N-2r+2}$. In this case, the highest exponent of x occurring in W_{ij} is strictly less than $d_{N-2r+1} + d_{N-2r+2} - 1$, unless $i = N-2r+1$ and $j = N-2r+2$, in which case $W_{ij} = W_{N-2r+2}$. It follows from the claim and the induction hypothesis that the term

$$P_{N-2r+1,N-2r+2}(\mathbf{x}_1,\dots,\mathbf{x}_{r-1})W_{N-2r+2}(\mathbf{x}_r)$$

has a monomial occurring in no other term of the sum (2.9), hence D_r is itself non-zero. The lemma now follows. □

[4] So $(a_1,\dots,a_n) \prec (b_1,\dots,b_n)$ if there exists i with $a_i < b_i$ and $a_j = b_j$ for all $j < i$.

The p-adic iterative method yields a congruence relation amongst the variables of equation (1.6). In order to use this relation to provide an iterative bound on $J_{s,\Phi}(X)$, we need to count the number of solutions to such a congruence. This is the purpose of the next lemma.

Definition 2.4. Given $\boldsymbol{\sigma} \in \{-1,1\}^M$, $\mathbf{m} \in \mathbb{Z}^N$, $\boldsymbol{\xi} \in \mathbb{Z}^2$ and a prime p, define $\mathcal{B}_p^{\boldsymbol{\sigma}}(\mathbf{m};\boldsymbol{\xi})$ to be the set of solutions $(\mathbf{x}_1,\dots,\mathbf{x}_M)$ modulo p^k of the system of equations

$$\sum_{j=1}^{M} \sigma_j F_i(\mathbf{x}_j - \boldsymbol{\xi}) \equiv m_i \pmod{p^{k_i}} \quad (1 \le i \le N) \tag{2.10}$$

satisfying the additional condition that $\Delta(\mathbf{x}_1,\dots,\mathbf{x}_M) \not\equiv 0 \bmod p$.

Lemma 2.5. *We have the upper bound*

$$|\mathcal{B}_p^{\boldsymbol{\sigma}}(\mathbf{m};\boldsymbol{\xi})| \le k_1 \cdots k_N \; p^{2Mk-K}. \tag{2.11}$$

In order to prove Lemma 2.5, we record a simple generalisation of Lagrange's theorem on the number of roots of a non-zero polynomial over an arbitrary field, a result which proves useful elsewhere.

Lemma 2.6. *Let $\mathbb{F}$ be a field and $P \in \mathbb{F}[X_1,\dots,X_m]$ a non-zero polynomial. Let a_i denote the highest exponent of X_i occurring in P. If $A \subset \mathbb{F}$ is finite then*

$$\#\left\{\mathbf{x} \in A^m : P(\mathbf{x}) = 0\right\} \le (a_1 + \cdots + a_m)|A|^{m-1}.$$

The proof is a simple induction on the number of variables m, which we leave as an exercise for the reader.

Proof of Lemma 2.5. Let $\mathcal{D}(\mathbf{n})$ denote the number of elements $(\mathbf{x}_1,\dots,\mathbf{x}_M)$ in the set $\mathcal{B}_p^{\boldsymbol{\sigma}}(\mathbf{m};\boldsymbol{\xi})$ satisfying the stronger congruence

$$\sum_{j=1}^{M} \sigma_j \mathbf{F}(\mathbf{x}_j - \boldsymbol{\xi}) \equiv \mathbf{n} \pmod{p^k}.$$

Then

$$\begin{aligned} |\mathcal{B}_p^{\boldsymbol{\sigma}}(\mathbf{m};\boldsymbol{\xi})| &\le \sum_{\substack{1 \le n_1 \le p^k \\ n_1 \equiv m_1 \bmod p^{k_1}}} \cdots \sum_{\substack{1 \le n_N \le p^k \\ n_N \equiv m_N \bmod p^{k_N}}} \mathcal{D}(\mathbf{n}) \\ &\le p^{(kN-K)} \max_{\mathbf{n}} \mathcal{D}(\mathbf{n}). \end{aligned} \tag{2.12}$$

For each tuple $(x_1, x_2, \dots, x_{2M-1}, x_{2M})$ counted by $\mathcal{D}(\mathbf{n})$ there are at most $p^{(2M-N)k}$ choices for x_i with $i > N$. Fix such a choice, and define the polynomials

$$f_i(x_1,\dots,x_N) = \sum_{j=1}^{M} \sigma_j F_i(x_{2j-1} - \xi_1, x_{2j} - \xi_2) - n_i \quad (1 \le i \le N).$$

Then $f_1,\dots,f_N$ are polynomials in $\mathbb{Z}[x_1,\dots,x_N]$ with $\deg f_i = k_i$. By Theorem 1 of Wooley [22], the number of integer tuples $1 \le (x_1,\dots,x_N) \le p^k$ satisfying both

$$f_i(x_1,\dots,x_N) \equiv 0 \mod p^k \qquad (1 \le i \le N)$$

and

$$\det\Big(\frac{\partial f_i}{\partial x_j}(\mathbf{x})\Big)_{i,j} \not\equiv 0 \mod p$$

is at most $(\deg f_1)\cdots(\deg f_N) = k_1\cdots k_N$. One can check using (1.11), that for $(u,v)=(1,0)$ or $(0,1)$, we have

$$F_i^{u,v}(\mathbf{x}_j - \boldsymbol{\xi}) = F_i^{u,v}(\mathbf{x}_j) + \sum_{l>i}(\Xi_{-\boldsymbol{\xi}})_{il}F_l^{u,v}(\mathbf{x}_j).$$

Hence it follows that

$$\left|\det\Big(\frac{\partial f_i}{\partial x_j}(x_1,\dots,x_N)\Big)_{i,j}\right| = \big|\Delta(x_1,x_2,\dots,x_{2M-1},x_{2M})\big|.$$

So there are at most $k_1\cdots k_N$ choices for $(x_1,\dots,x_N)$ with $(x_1,\dots,x_{2M})$ counted by $\mathcal{D}(\mathbf{n})$. Thus

$$\mathcal{D}(\mathbf{n}) \le k_1\cdots k_N p^{(2M-N)k}. \tag{2.13}$$

Putting (2.12) and (2.13) together, we obtain the lemma. □

Lemma 2.5 allows us to count non-singular solutions, which we have still to define. The remaining singular solutions are counted by the following lemma. First a definition.

Definition 2.7. Define $\mathcal{S}_t(X)$ to be the set of

$$(\mathbf{x}_1,\dots,\mathbf{x}_t)\in[X]^{2t}$$

such that for any function $h:[M]\to[t]$ we have the identity

$$\Delta(\mathbf{x}_{h(1)},\dots,\mathbf{x}_{h(M)}) = 0.$$

Lemma 2.8. *Let Φ be a non-degenerate binary form of degree k and differential dimension N. Setting $M=\lceil N/2\rceil$, we have the upper bound*

$$|\mathcal{S}_t(X)| \le Mt^M(2k)^t X^{t+M-1}. \tag{2.14}$$

Proof. The result follows trivially if $t<M$, so we may assume that $t\ge M$. Let us define a sequence of non-zero polynomials $\Delta_i(\mathbf{x}_1,\dots,\mathbf{x}_i)$ for $i=0,1,\dots,M$. We begin by setting $\Delta_M=\Delta$. Suppose we have constructed Δ_i with $i>1$. Let us write $\mathbf{x}^{\mathbf{a}}$ for the monomial $x_1^{a_1}x_2^{a_2}$. Of the monomials $\mathbf{x}_1^{\mathbf{a}_1}\cdots\mathbf{x}_i^{\mathbf{a}_i}$ occurring in Δ_i, let $\mathbf{b}_i$ denote the maximum in the lexicographical ordering over all $\mathbf{a}_i$. It follows that there exist polynomials $\Delta_{i-1}(\mathbf{x}_1,\dots,\mathbf{x}_{i-1})$ and $R_i(\mathbf{x}_1,\dots,\mathbf{x}_i)$ such that

$$\Delta_i(\mathbf{x}_1,\dots,\mathbf{x}_i) = \Delta_{i-1}(\mathbf{x}_1,\dots,\mathbf{x}_{i-1})\mathbf{x}_i^{\mathbf{b}_i} + R_i(\mathbf{x}_1,\dots,\mathbf{x}_i). \tag{2.15}$$

Moreover, we may assume Δ_{i-1} is non-zero and that every monomial $\mathbf{x}_1^{\mathbf{a}_1}\cdots\mathbf{x}_i^{\mathbf{a}_i}$ occurring in R_i satisfies $\mathbf{a}_i\prec\mathbf{b}_i$, where $\prec$ denotes the (strict) lexicographical ordering. For consistency, let us set $\Delta_0=1$ and $R_1=0$. For each i in the range $1\le i\le M$, define $\mathcal{T}_i$ to be the set of $(\mathbf{x}_1,\dots,\mathbf{x}_t)\in[X]^{2t}$ satisfying both of the following conditions:

(i) For any $h:[i]\to[t]$ we have

$$\Delta_i(\mathbf{x}_{h(1)},\dots,\mathbf{x}_{h(i)}) = 0,$$

(ii) For each $j<i$ there exists $h:[j]\to[t]$ such that

$$\Delta_j(\mathbf{x}_{h(1)},\dots,\mathbf{x}_{h(j)}) \ne 0.$$

Then we have that
$$\mathcal{S}_t(X) \subset \bigcup_{1\le i\le M} \mathcal{T}_i.$$

Let $(\mathbf{x}_1,\dots,\mathbf{x}_M) \in \mathcal{T}_i$. Then there exists some $h : [i-1] \to [t]$ such that
$$\Delta_{i-1}(\mathbf{x}_{h(1)},\dots,\mathbf{x}_{h(i-1)}) \neq 0,$$
yet for all $j \notin \{h(1),\dots,h(i-1)\}$, the identity (2.15) tells us that
$$0 = \Delta_{i-1}(\mathbf{x}_{h(1)},\dots,\mathbf{x}_{h(i-1)})\mathbf{x}_j^{\mathbf{b}_j} + R_j(\mathbf{x}_{h(1)},\dots,\mathbf{x}_{h(i-1)},\mathbf{x}_j). \tag{2.16}$$
Since the two-variable polynomial
$$Q(\mathbf{X}) = \Delta_{i-1}(\mathbf{x}_{h(1)},\dots,\mathbf{x}_{h(i-1)})\mathbf{X}^{\mathbf{b}_j} + R_j(\mathbf{x}_{h(1)},\dots,\mathbf{x}_{h(i-1)},\mathbf{X})$$
is non-zero, it follows from Lemma 2.6 that for each j, the number of $\mathbf{x}_j$ satisfying (2.16) is at most
$$(b_{j1}+b_{j2})X \le 2kX.$$
There are trivially at most $X^{2(i-1)}$ choices for $(\mathbf{x}_{h(1)},\dots,\mathbf{x}_{h(i-1)})$, and at most t^{i-1} choices for $h : [i-1] \to [t]$. Thus
$$|\mathcal{T}_i| \le t^{i-1}(2k)^{t-i+1}X^{2(i-1)+(t-i+1)} = t^M(2k)^tX^{t+i-1}.$$
Hence
$$|\mathcal{S}_t(X)| \le Mt^M(2k)^tX^{t+M-1}.$$
□

We can now implement the results obtained so far in this section to prove the following lemma, which encodes the basic iterative relation underlying our p-adic approach to bounding $J_{s,\Phi}$. Again, we begin with a definition.

Definition 2.9. Given a prime number p, $\boldsymbol{\xi} \in \mathbb{Z}^2$ and $\boldsymbol{\sigma} \in \{-1,1\}^M$, define the exponential sums
$$\mathfrak{f}_p(\boldsymbol{\alpha};\boldsymbol{\xi}) = \sum_{\substack{\mathbf{x}\in[X]^2\\ \mathbf{x}\equiv\boldsymbol{\xi} \bmod p}} e(\boldsymbol{\alpha}\cdot\mathbf{F}(\mathbf{x})),$$
$$\mathfrak{F}_p^{\boldsymbol{\sigma}}(\boldsymbol{\alpha}) = \sum_{\substack{(\mathbf{x}_1,\dots,\mathbf{x}_M)\in[X]^{2M}\\ \Delta(\mathbf{x}_1,\dots,\mathbf{x}_M)\not\equiv 0 \bmod p}} e\Big(\boldsymbol{\alpha}\cdot\sum_{j=1}^M \sigma_j\mathbf{F}(\mathbf{x}_j)\Big).$$

Lemma 2.10. *Let $s \ge M$. Then there exists $\boldsymbol{\xi} \in \mathbb{Z}^2$, $\boldsymbol{\sigma} \in \{-1,1\}^M$ and a prime p in the range $X^{1/k} < p \le 2X^{1/k}$ such that*
$$J_{s,\Phi}(X) \ll X^{2s+M-1} + p^{4(s-M)}\oint |\mathfrak{F}_p^{\boldsymbol{\sigma}}(\boldsymbol{\alpha})|^2|\mathfrak{f}_p(\boldsymbol{\alpha};\boldsymbol{\xi})|^{2(s-M)}\mathrm{d}\boldsymbol{\alpha}. \tag{2.17}$$

Proof. If $(\mathbf{x}_1,\dots,\mathbf{x}_M) \in [X]^{2M}$, then each ij-entry of the matrix $\mathrm{Jac}(\mathbf{x}_1,\dots,\mathbf{x}_M)$, defined in (2.2), is of order $O(X^{k_i-1})$. Hence there exists a constant $C = C(\Phi)$ such that for any $(\mathbf{x}_1,\dots,\mathbf{x}_N) \in [X]^{2M}$ we have
$$|\Delta(\mathbf{x}_1,\dots,\mathbf{x}_M)| \le CX^{K-N}.$$
By the prime number theorem, for all $X \gg_\Phi 1$ we have
$$\pi(2X^{1/k}) - \pi(X^{1/k}) \ge \frac{k\log C}{\log X} + k(K-N). \tag{2.18}$$

Let T be the smallest positive integer bounded below by the right-hand side of (2.18), and let $\mathcal{P}$ denote the set of the T smallest primes in the interval $X^{1/k} < p \leq 2X^{1/k}$. Then

$$\prod_{p\in\mathcal{P}} p > X^{T/k} \geq CX^{K-N}.$$

In particular, for each $(\mathbf{x}_1,\ldots,\mathbf{x}_M) \in [X]^{2M}$ with $\Delta(\mathbf{x}_1,\ldots,\mathbf{x}_M) \neq 0$, there must exist $p \in \mathcal{P}$ such that

$$\Delta(\mathbf{x}_1,\ldots,\mathbf{x}_M) \not\equiv 0 \pmod{p}.$$

Now $J_{s,\Phi}(X)$ counts tuples $(\mathbf{x}_1,\ldots,\mathbf{x}_{2s}) \in [X]^{4s}$ which satisfy

$$\sum_{i=1}^{s}(\mathbf{F}(\mathbf{x}_{2i-1}) - \mathbf{F}(\mathbf{x}_{2i})) = 0. \tag{2.19}$$

Let $\mathcal{T}$ denote the number of such tuples which are not contained in $\mathcal{S}_{2s}(X)$. Then by Lemma 2.8 we have

$$J_{s,\Phi}(X) \ll_{s,\Phi} X^{2s+M-1} + \mathcal{T}.$$

If $(\mathbf{x}_1,\ldots,\mathbf{x}_{2s})$ denotes a tuple counted by $\mathcal{T}$, then it satisfies (2.19) and there exists $h : [M] \to [2s]$ such that $\Delta(\mathbf{x}_{h(1)},\ldots,\mathbf{x}_{h(M)}) \neq 0$. Hence for each such choice of tuple and function h, there exists a prime $p \in \mathcal{P}$ such that

$$\Delta(\mathbf{x}_{h(1)},\ldots,\mathbf{x}_{h(M)}) \not\equiv 0 \pmod{p}.$$

Notice that by the definition of the determinant Δ, such a choice of h must be injective. Since there are $O(1)$ choices for h, and $O(1)$ choices for a prime $p \in \mathcal{P}$, we see that there exists p and $\boldsymbol{\sigma} \in \{-1,1\}^M$ such that

$$\begin{aligned}\mathcal{T} &\ll \oint |\mathfrak{F}_p^{\boldsymbol{\sigma}}(\boldsymbol{\alpha})||f(\boldsymbol{\alpha})|^{2s-M}\mathrm{d}\boldsymbol{\alpha}\\ &\leq \Big(\oint |\mathfrak{F}_p^{\boldsymbol{\sigma}}(\boldsymbol{\alpha})|^2|f(\boldsymbol{\alpha})|^{2(s-M)}\mathrm{d}\boldsymbol{\alpha}\Big)^{1/2} J_{s,\Phi}(X)^{1/2}.\end{aligned}$$

Thus

$$J_{s,\Phi}(X) \ll X^{2s+M-1} + \oint |\mathfrak{F}_p^{\boldsymbol{\sigma}}(\boldsymbol{\alpha})|^2|f(\boldsymbol{\alpha})|^{2(s-M)}\mathrm{d}\boldsymbol{\alpha}. \tag{2.20}$$

By the triangle inequality

$$|f(\boldsymbol{\alpha})|^{2(s-M)} = \Big|\sum_{\boldsymbol{\xi}\in[p]^2} \mathfrak{f}_p(\boldsymbol{\alpha};\boldsymbol{\xi})\Big|^{2(s-M)} \leq p^{4(s-M)} \max_{\boldsymbol{\xi}} |\mathfrak{f}_p(\boldsymbol{\alpha};\boldsymbol{\xi})|^{2(s-M)}.$$

Incorporating this into (2.20), we obtain the lemma. □

The following lemma eventually allows us to conclude that the first term on the right-hand side of (2.17) is smaller than our hoped for upper bound.

Lemma 2.11. *Let Φ be a non-degenerate binary form of degree k, differential dimension N and differential degree K. Set $M = \lceil N/2 \rceil$. Then we have the inequality*

$$K/k \leq M + \tfrac{1}{2}. \tag{2.21}$$

Proof. We establish this result in a series of claims. Let N_l denote the size of the set $\{i : k_i = l\}$.

Claim 1. *Let $2 \leq l \leq k$. Then $N_{l-1} \geq N_l$ or $N_l = l + 1$.*

Let $G_1, \ldots, G_m$ denote a basis of forms for the space $\operatorname{span}\{F_i : k_i = l\}$, so that $m = N_l$. Let d_j denote the degree of the one variable polynomial $G_j(x,1)$. By performing a linear transformation we may assume that $d_1 > d_2 > \cdots > d_m$.

Suppose there exists i for which $d_i < l + 1 - i$. Let i denote the minimal such index. Then each of the one variable polynomials $G_j^{1,0}(x,1)$ with $1 \leq j < i$ has degree $d_j - 1$, whilst each of the polynomials $G_j^{0,1}(x,1)$ with $j \geq i$ has degree d_j. Since

$$\begin{aligned} d_1 - 1 > d_2 - 1 > \cdots > d_{i-1} - 1 &= l + 1 - i \\ &> d_i > \cdots > d_m, \end{aligned}$$

the polynomials $G_1^{1,0}, \ldots, G_{i-1}^{1,0}, G_i^{0,1}, \ldots, G_m^{0,1}$ are a linearly independent subset of the space $\operatorname{span}\{F_i : k_i = l - 1\}$. It follows that $N_{l-1} \geq N_l$.

Next suppose that for all i we have $d_i = l + i - 1$. If $N_l < l + 1$ then, as above, $G_1^{1,0}, \ldots, G_m^{1,0}$ form a linearly independent subset of $\operatorname{span}\{F_i : k_i = l - 1\}$ of size N_l and we are done. The only remaining possibility is that $N_l = l + 1$. This establishes Claim 1.

Claim 2. *For $2 \leq l \leq k$ we have the inequality*

$$N_{l-1} \leq N_l + 1.$$

Let $G_1, \ldots, G_m$ denote a basis for $\operatorname{span}\{F_i : k_i = l\}$. Then for each u, v with $l = k - u - v$, the form $\Phi^{u,v}$ is a linear combination of $G_1, \ldots, G_m$. It follows that each $\Phi^{u+1,v}$ is a linear combination of $G_1^{1,0}, \ldots, G_m^{1,0}$. If $u > 0$ then $\Phi^{u,v+1}$ is also a linear combination of the $G_i^{1,0}$, since $(u, v+1) = (u'+1, v')$ for some $u', v' \geq 0$ and $l = d - u' - v'$. Thus

$$\operatorname{span}\{F_i : k_i = l - 1\} = \operatorname{span}\left\{G_1^{1,0}, \ldots, G_m^{1,0},\ \Phi^{0,d-l+1}\right\}.$$

Clearly this latter space has dimension at most $N_l + 1$, which is what we require.

Claim 3. *If $1 \leq l \leq k/2$ then $N_{k-l} \leq N_l$.*

Let $1 \leq l < k/2$. If $N_i < i + 1$ for all $l < i \leq k - l$ then by Claim 1 we have

$$N_{k-l} \leq \cdots \leq N_{l+1} \leq N_l.$$

Next suppose $N_i = i+1$ for some $l < i \leq k-l$. Since $\operatorname{span}\{F_j : k_j = i\}$ is a subspace of the $(i+1)$-dimensional space

$$\operatorname{span}\left\{x^{i-j}y^j : 0 \leq j \leq i\right\},$$

these spaces must in fact coincide. Taking derivatives, we see that $\operatorname{span}\{F_j : k_j = l\}$ coincides with $\operatorname{span}\left\{x^{l-j}y^j : 0 \leq j \leq l\right\}$, so $N_l = l + 1$. By Claim 2, we have

$$N_{k-l} \leq N_{k-(l-1)} + 1 \leq \cdots \leq N_k + l \leq 1 + l = N_l,$$

the last inequality being a consequence of $N_k = 1$. This establishes Claim 3.

Finally, we use Claim 3 to prove the lemma. Observing that $\frac{1}{2} + M \geq (N+1)/2$, the inequality (2.21) therefore reduces to showing that

$$\sum_{l=1}^{k} l N_l \leq \tfrac{k}{2} + \sum_{l=1}^{k} \tfrac{k}{2} N_l.$$

Re-arranging, we need only show

$$\sum_{k/2<l\leq k}\left(l-\tfrac{k}{2}\right)N_l \leq \tfrac{k}{2}+\sum_{1\leq l<k/2}\left(\tfrac{k}{2}-l\right)N_l.$$

Changing variables from l to $k-l$ on the left hand side leaves us the task of proving

$$\tfrac{k}{2}N_k+\sum_{1\leq l<k/2}\left(\tfrac{k}{2}-l\right)N_{k-l} \leq \tfrac{k}{2}+\sum_{1\leq l<k/2}\left(\tfrac{k}{2}-l\right)N_l.$$

This last inequality follows from Lemma 2.2 and the fact that $N_k = 1$. □

The final lemma proved before we deduce Theorem 1.7 bounds the second term on the right-hand side of (2.17).

Lemma 2.12. *Suppose that $s \geq M$ and $X^{1/k} \leq p \leq X$. Then*

$$\oint |\mathfrak{F}_p^{\boldsymbol{\sigma}}(\boldsymbol{\alpha})|^2|\mathfrak{f}_p(\boldsymbol{\alpha};\boldsymbol{\xi})|^{2(s-M)}\mathrm{d}\boldsymbol{\alpha} \ll X^{2M}p^{2Mk-K}J_{s-M,\Phi}(2X/p). \tag{2.22}$$

Proof. The left-hand side of (2.22) counts tuples $(\mathbf{x}_1,\mathbf{y}_1,\ldots,\mathbf{x}_s,\mathbf{y}_s) \in [X]^{4s}$ satisfying the Diophantine equations

$$\sum_{j=1}^{M}\sigma_j\big(\mathbf{F}(\mathbf{x}_j)-\mathbf{F}(\mathbf{y}_j)\big) = \sum_{j=M+1}^{s}\big(\mathbf{F}(\mathbf{x}_j)-\mathbf{F}(\mathbf{y}_j)\big), \tag{2.23}$$

under the additional constraints that both $\Delta(\mathbf{x}_1,\ldots,\mathbf{x}_M)$ and $\Delta(\mathbf{y}_1,\ldots,\mathbf{y}_M)$ are non-zero modulo p, and for all $j > M$ we have $\mathbf{x}_j \equiv \mathbf{y}_j \equiv \boldsymbol{\xi} \pmod{p}$. Translation-invariance (1.12) and homogeneity together imply that

$$\sum_{j=1}^{M}\sigma_j\Big(F_i(\mathbf{x}_j-\boldsymbol{\xi})-F_i(\mathbf{y}_j-\boldsymbol{\xi})\Big) \equiv 0 \mod p^{k_i} \qquad (1\leq i\leq N).$$

Fix a choice of $(\mathbf{y}_1,\ldots,\mathbf{y}_M) \in [X]^{2M}$ and set

$$\mathbf{m} = \sum_{j=1}^{M}\sigma_j F_i(\mathbf{y}_j-\boldsymbol{\xi}).$$

Let $\overline{x}$ denote the residue class of $x \in \mathbb{Z}$ modulo p^k. Then $(\overline{\mathbf{x}}_1,\ldots,\overline{\mathbf{x}}_M) \in \mathcal{B}_p^{\boldsymbol{\sigma}}(\mathbf{m};\boldsymbol{\xi})$. Since $p^k \geq X$, the map

$$(\mathbf{x}_1,\ldots,\mathbf{x}_M) \mapsto (\overline{\mathbf{x}}_1,\ldots,\overline{\mathbf{x}}_M)$$

is injective when restricted to $[X]^{2M}$. Hence, there are at most $|\mathcal{B}_p^{\boldsymbol{\sigma}}(\mathbf{m};\boldsymbol{\xi})|$ choices for $(\mathbf{x}_1,\ldots,\mathbf{x}_M)$. Set

$$\mathbf{n} = \sum_{j=1}^{M}\sigma_j\Big(F_i(\mathbf{x}_j-\boldsymbol{\xi})-F_i(\mathbf{y}_j-\boldsymbol{\xi})\Big).$$

Then for each fixed choice of $(\mathbf{x}_1,\mathbf{y}_1,\ldots,\mathbf{x}_M,\mathbf{y}_M)$, the number of choices for the remaining $\mathbf{x}_j,\mathbf{y}_j$ $(j>M)$ is at most

$$\oint |\mathfrak{f}_p(\boldsymbol{\alpha};\boldsymbol{\xi})|^{2(s-M)}e(-\boldsymbol{\alpha}\cdot\mathbf{n})\mathrm{d}\boldsymbol{\alpha} \leq \oint |\mathfrak{f}_p(\boldsymbol{\alpha};\boldsymbol{\xi})|^{2(s-M)}\mathrm{d}\boldsymbol{\alpha}.$$

Employing Lemma 2.5, it remains to establish that

$$\oint |\mathfrak{f}_p(\boldsymbol{\alpha};\boldsymbol{\xi})|^{2t}\mathrm{d}\boldsymbol{\alpha} \ll_t J_t(2X/p). \tag{2.24}$$

We may assume $\boldsymbol{\xi} \in [p]^2$. The integral in (2.24) then counts the number of solutions to the system

$$\sum_{i=1}^{t}(\mathbf{F}(\boldsymbol{\xi}+p\mathbf{x}_i)-\mathbf{F}(\boldsymbol{\xi}+p\mathbf{y}_i))=0, \quad \mathbf{x}_i,\mathbf{y}_i \in \left[0,\tfrac{X-\xi_1}{p}\right]\times\left[0,\tfrac{X-\xi_2}{p}\right].$$

By (1.12) and homogeneity, this equals the number of solutions of the system

$$\sum_{i=1}^{t}(\mathbf{F}(\mathbf{x}_i)-\mathbf{F}(\mathbf{y}_i))=0, \quad \mathbf{x}_i,\mathbf{y}_i \in \left[1,\tfrac{X-\xi_1}{p}+1\right]\times\left[1,\tfrac{X-\xi_2}{p}+1\right].$$

Since we may assume $X \geq p$, the result follows. □

To conclude this section, we prove our mean value theorem.

Proof of the Theorem 1.7. We proceed by induction on $\lfloor s/M\rfloor \geq 0$. The basis case is equivalent to $J_{s,\Phi}(X) \ll X^{4s}$, which is trivial.

Let us suppose that $\lfloor s/M\rfloor \geq 1$. Combining Lemma 2.10 and Lemma 2.12, we see that there exists a prime p in the interval $(X^{1/k}, 2X^{1/k}]$ such that

$$J_{s,\Phi}(X) \ll X^{2s+M-1} + p^{4(s-M)-K}X^{4M}J_{s-M,\Phi}(2X/p). \tag{2.25}$$

Combining this with our induction hypothesis implies that

$$J_{s,\Phi}(X) \ll X^{2s+M-1} + X^{4s-K+\Delta_s}.$$

It therefore remains to show that $2s+M-1 \leq 4s-K+\Delta_s$, which we also prove by induction on $\lfloor s/M\rfloor \geq 1$. The basis case follows directly from the estimate $K/k \leq M+\frac{1}{2}$ of Lemma 2.11. Let us suppose $\lfloor s/M\rfloor \geq 2$, then by induction $2s+M-1$, being equal to $2M+2(s-M)+M-1$, is at most

$$\begin{aligned}2M+4(s-M)-K+K(1-\tfrac{1}{k})^{\lfloor \frac{s}{M}\rfloor-1} &= 4s-K+\Delta_s+\tfrac{K}{k}(1-\tfrac{1}{k})^{\lfloor\frac{s}{M}\rfloor-1}-2M\\ &\leq 4s-K+\Delta_s+\tfrac{K}{k}-2M.\end{aligned}$$

Since $K=\sum_{i=1}^{N}k_i \leq Nk$ and $N \leq 2M$, we have $\frac{K}{k}-2M \leq 0$, which completes the proof. □

3. Weyl-type estimates

Definition 3.1. Let us say $\Delta = \Delta_s$ is an *admissible exponent* for (s,Φ) if there exists a constant $C = C(s,\Phi)$ such that for any $X \geq 1$ we have the bound $J_{s,\Phi}(X) \leq CX^{4s-K+\Delta}$.

The aim of this section is to prove the following Weyl-type estimate.

Theorem 3.2. *Let Δ be an admissible exponent for (s,Φ) and let $\sigma < \frac{1-3\Delta}{6s+3}$. Then for any $\varepsilon > 0$ there exists a constant $C = C(s,\Phi,\varepsilon)$ such that if $X \geq C$ and*

$$|f(\boldsymbol{\alpha};X)| \geq X^{2-\sigma}, \tag{3.1}$$

then there exists integers $q, a_1, \ldots, a_N$ such that $1 \leq q \leq X^{k\sigma+\varepsilon}$ and $|q\alpha_j - a_j| \leq X^{k\sigma+\varepsilon-k_j}$ for all $1 \leq j \leq N$.

The proof of Theorem 3.2 uses Vinogradov's method, a general heuristic for which is neatly described in [8, §8.5]. We model our argument on a version of the method due to Vaughan [21, Chapter 5], with a later improvement due to Baker [2, Chapter 4].

Let $\gamma_{ij}(\boldsymbol{\xi})$ denote the ij-entry of the matrix $\Xi_{\boldsymbol{\xi}}$ occurring in (1.11). Then for any $\mathbf{y}$ we have the identity

$$F_i(\mathbf{x}+\mathbf{y}) - F_i(\mathbf{y}) = \sum_j \gamma_{ij}(\mathbf{y}) F_j(\mathbf{x}) = F_i(\mathbf{x}) + \sum_{j>i} \gamma_{ij}(\mathbf{y}) F_j(\mathbf{x}).$$

Set

$$\gamma_j(\mathbf{y}) = \gamma_j(\mathbf{y};\boldsymbol{\alpha}) = \sum_{i<j} \alpha_i \gamma_{ij}(\mathbf{y}),$$

then for all $\mathbf{x}$ we have

$$\boldsymbol{\alpha}\cdot\Big(\mathbf{F}(\mathbf{x}+\mathbf{y}) - \mathbf{F}(\mathbf{x}) - \mathbf{F}(\mathbf{y})\Big) = \sum_{j=2}^{N} \gamma_j(\mathbf{y}) F_j(\mathbf{x}). \tag{3.2}$$

In the following result, and the remainder of the paper, we use $\|\beta\|$ to denote the smallest distance from β to an integer.

Lemma 3.3. *Let Δ denote an admissible exponent for (s,Φ) and let $\mathcal{S}$ be a subset of $[X]^2$ of size S such that for any distinct $\mathbf{y},\mathbf{z}\in\mathcal{S}$ there exists $2\le j\le N$ with*

$$\|\gamma_j(\mathbf{y}) - \gamma_j(\mathbf{z})\| > X^{-k_j}. \tag{3.3}$$

Then one has

$$|f(\boldsymbol{\alpha};X)| \ll X^2 \log(2X)^2 \left(X^{\Delta}/S\right)^{1/(2s)}. \tag{3.4}$$

Proof. Averaging, we see that $f(\boldsymbol{\alpha})$ is equal to

$$\frac{1}{S}\sum_{\mathbf{y}\in\mathcal{S}} \sum_{\mathbf{x}\in(-X,X)^2} e\big(\boldsymbol{\alpha}\cdot\mathbf{F}(\mathbf{x}+\mathbf{y})\big) 1_{[X]^2-\mathbf{y}}(\mathbf{x}). \tag{3.5}$$

By orthogonality

$$1_{[X]^2-\mathbf{y}}(\mathbf{x}) = \oint e(\boldsymbol{\beta}\cdot\mathbf{x}) \sum_{\mathbf{n}\in[X]^2-\mathbf{y}} e(-\boldsymbol{\beta}\cdot\mathbf{n})\mathrm{d}\boldsymbol{\beta}.$$

Since $\sum_{1-y\le n\le X-y} e(-\beta n) \le \min\left\{X, \|\beta\|^{-1}\right\}$, the sum $f(\boldsymbol{\alpha})$ is at most

$$\begin{aligned}&\frac{1}{S}\oint\Big| \sum_{\substack{\mathbf{y}\in\mathcal{S}\\ \mathbf{x}\in(-X,X)^2}} e\big(\boldsymbol{\alpha}\cdot\mathbf{F}(\mathbf{x}+\mathbf{y}) + \boldsymbol{\beta}\cdot\mathbf{x}\big)\Big| \min\left\{X, \|\beta_1\|^{-1}\right\}\min\left\{X, \|\beta_2\|^{-1}\right\}\mathrm{d}\boldsymbol{\beta}\\ &\ll \frac{\log(2X)^2}{S}\sup_{\boldsymbol{\beta}} \sum_{\mathbf{y}\in\mathcal{S}} |g(\mathbf{y},\boldsymbol{\beta})|,\end{aligned}$$

where

$$g(\mathbf{y},\boldsymbol{\beta}) = \sum_{\mathbf{x}\in(-X,X)^2} e\left(\boldsymbol{\alpha}\cdot\mathbf{F}(\mathbf{x}+\mathbf{y}) + \boldsymbol{\beta}\cdot\mathbf{x}\right).$$

It follows from Hölder's inequality that there exists $\boldsymbol{\beta}\in\mathbb{T}^2$ such that

$$|f(\boldsymbol{\alpha})|^{2s} \ll \frac{\log(2X)^{4s}}{S}\sum_{\mathbf{y}\in\mathcal{S}} |g(\mathbf{y},\boldsymbol{\beta})|^{2s}. \tag{3.6}$$

Let $C = C(s, \Phi)$ be any constant such that for all $1 \le i \le N$ we have

$$\Big|\sum_{j=1}^{s} F_i(\mathbf{x}_j)\Big| \le CX^{k_i} \quad (\mathbf{x}_j \in (-X, X)^2).$$

Define $\mathbf{F}'(\mathbf{x}) = (F_i(\mathbf{x}))_{2 \le i \le N}$, $\mathcal{N} = \prod_{2 \le i \le N}[-CX^{k_i}, CX^{k_i}]$ and

$$a(\mathbf{n}) = \sum_{\substack{\mathbf{x}_1, \dots, \mathbf{x}_s \in (-X,X)^2 \\ \mathbf{n} = \sum_{j=1}^{s} \mathbf{F}'(\mathbf{x}_j)}} e\Big(\boldsymbol{\alpha} \cdot \sum_{j=1}^{s} \mathbf{F}(\mathbf{x}_j) + \boldsymbol{\beta} \cdot \sum_{j=1}^{s} \mathbf{x}_j\Big).$$

Then by (3.2) we have

$$\begin{aligned}\Big|\sum_{\mathbf{x} \in (-X,X)^2} & e(\boldsymbol{\alpha} \cdot \mathbf{F}(\mathbf{x} + \mathbf{y}) + \boldsymbol{\beta} \cdot \mathbf{x})\Big|^{2s} \\ &= \Big|\sum_{\mathbf{x}_1, \dots, \mathbf{x}_s \in (-X,X)^2} e\Big(\boldsymbol{\alpha} \cdot \sum_{j=1}^{s} \mathbf{F}(\mathbf{x}_j) + \boldsymbol{\beta} \cdot \sum_{j=1}^{s} \mathbf{x}_j\Big) e\Big(\boldsymbol{\gamma}(\mathbf{y}) \cdot \sum_{j=1}^{s} \mathbf{F}'(\mathbf{x}_j)\Big)\Big|^2 \\ &= \Big|\sum_{\mathbf{n} \in \mathcal{N}} a(\mathbf{n}) e(\boldsymbol{\gamma}(\mathbf{y}) \cdot \mathbf{n})\Big|^2.\end{aligned}$$

By (3.3) and the multi-dimensional version of the large sieve inequality (see for example Vaughan [21, Lemma 5.3]), we have

$$\sum_{\mathbf{y} \in \mathcal{S}} \Big|\sum_{\mathbf{n} \in \mathcal{N}} a(\mathbf{n}) e(\boldsymbol{\gamma}(\mathbf{y}) \cdot \mathbf{n})\Big|^2 \ll_{s,\Phi} \sum_{\mathbf{n} \in \mathcal{N}} |a(\mathbf{n})|^2 \prod_{2 \le i \le N} X^{k_i}. \tag{3.7}$$

Let $\mathbf{e}_1$ denote the first standard basis vector. Recalling that $J_{s,\Phi}(X; \mathbf{m})$ denotes the number of $(\mathbf{x}, \mathbf{y}) \in [X]^{4s}$ satisfying

$$\sum_{i=1}^{s} \big(\mathbf{F}(\mathbf{x}_i) - \mathbf{F}(\mathbf{y}_i)\big) = \mathbf{m},$$

we have, by translation invariance, that

$$\sum_{\mathcal{N}} |a(\mathbf{n})|^2 \le \sum_{m} J_{s,\Phi}(2X + 1; m\mathbf{e}_1), \tag{3.8}$$

where the summation over m ranges over a set of size CX^{k_1}. If Δ is an admissible exponent for (s, Φ) then we deduce that the right-hand side of (3.8) is of order $O(X^{4s-(K-k_1)+\Delta})$. Putting these facts together with (3.7), we obtain

$$|f(\boldsymbol{\alpha})|^{2s} \ll_{s,\Phi} \frac{\log(2X)^{4s}}{S} X^{4s+\Delta}.$$

□

Remark 3.4. In the proof of Lemma 3.3, we used $J_{s,\Phi}(X)$ to bound the number $J'_{s,\Phi}(X)$ of solutions $(\mathbf{x}, \mathbf{y}) \in [X]^{4s}$ to the smaller system of equations

$$\sum_{i=1}^{s} \big(\Phi^{u,v}(\mathbf{x}_i) - \Phi^{u,v}(\mathbf{y}_i)\big) = 0 \qquad (u + v \ge 1).$$

We note that the methods of §2 translate almost verbatim to yield the bound $J'_{s,\Phi}(X) \ll X^{4s-(K-k)+\Delta}$, where for $s = l\,\lceil (N-1)/2 \rceil$ we have

$$\Delta \le (K - k)(1 - \tfrac{1}{k-1})^l.$$

This Δ is clearly superior to that obtained using $J_{s,\Phi}(X)$. However, at the level of detail we are concerned with, this makes little difference to our final results, and increases the expositional complexity of §2 considerably.

In order to obtain a set $\mathcal{S}$ satisfying the spacing condition (3.3), we relate this condition to the Diophantine approximation of our original coefficients α_i. This is the content of the following lemma. Unfortunately, the most direct approach allows us to control the spacing of the $\gamma_j(\mathbf{y})$ only according to the Diophantine approximation of a proper subset of the α_i. We first define this subset.

Definition 3.5. We assume throughout that I_1, I_2 denote sets of indices whose union equals $\{r \in [N] : k_r \geq 2\}$ and such that if $\{i, j\} = \{1, 2\}$ then both of the following conditions hold

$$\operatorname{span}\{F_r(x_1, x_2) : r \in I_i\} \cap \mathbb{Q}[x_j] = \{0\}, \tag{3.9}$$

$$\{F_r(x_1, x_2) : r \notin I_i\} \subset \mathbb{Q}[x_j]. \tag{3.10}$$

Making a linear transformation of the F_i if necessary, we can always guarantee the existence of such I_1 and I_2.

Lemma 3.6. *Let $m \in \{1, 2\}$. There exists an absolute constant $C = C(\Phi)$ and a positive integer $L \leq C$ such that for any $y, z \in [X]$ and $\boldsymbol{\alpha} \in \mathbb{T}^N$, if*

$$\|\gamma_j(y\mathbf{e}_m) - \gamma_j(z\mathbf{e}_m)\| \leq X^{-k_j} \quad \textit{for all } 2 \leq j \leq N, \tag{3.11}$$

then

$$\|L\alpha_i(y - z)\| \leq CX^{1-k_i} \quad \textit{for all } i \in I_m. \tag{3.12}$$

Proof. Let us suppose that $m = 1$, the case $m = 2$ being similar. By Taylor's formula, we have

$$F_i(\mathbf{x} + y\mathbf{e}_1) = F_i(\mathbf{x}) + \sum_{r=1}^{k_i-1} \frac{y^r}{r!} F_i^{r,0}(\mathbf{x}) + F_i(y\mathbf{e}_1). \tag{3.13}$$

Since $F_1, \ldots, F_N$ are a spanning subset of $\{\Phi^{u,v} : 0 \leq u + v < k\}$, there must exist rationals $\lambda_{ij}^{r,0}$ such that for $1 \leq r < k_i$ we have

$$F_i^{r,0} = \sum_{\substack{j \\ k_j = k_i - r}} \lambda_{ij}^{r,0} F_j. \tag{3.14}$$

Combining (3.13), (3.14) and (3.2), we obtain

$$\begin{aligned} \gamma_j(y\mathbf{e}_1) = \gamma_j(y\mathbf{e}_1; \boldsymbol{\alpha}) &= \sum_{r=1}^{k-k_j} \frac{y^r}{r!} \Bigg(\sum_{\substack{i \\ k_i = k_j + r}} \alpha_i \lambda_{ij}^{r,0} \Bigg) \\ &= \sum_{r=1}^{k-k_j} \frac{y^r}{r!} \gamma_j^{(r)}, \text{ say.} \end{aligned}$$

Notice that for $r \geq 2$ we have

$$\sum_j \lambda_{ij}^{r,0} F_j = F_i^{r,0} = \sum_t \lambda_{it}^{1,0} F_t^{r-1,0} = \sum_{t,j} \lambda_{it}^{1,0} \lambda_{tj}^{r-1,0} F_j.$$

Hence, by linear independence, for all i, j and $r \geq 2$ we have $\lambda_{ij}^{r,0} = \sum_t \lambda_{it}^{r-1,0} \lambda_{tj}^{1,0}$. It follows that

$$\gamma_j(y\mathbf{e}_1) = y\gamma_j^{(1)} + \sum_{r \geq 2} \frac{y^r}{r!} \sum_{\substack{t \\ k_t = k_j + r - 1}} \lambda_{tj}^{r-1,0} \Bigg(\sum_{\substack{i \\ k_i = k_t + 1}} \alpha_i \lambda_{it}^{1,0} \Bigg), \tag{3.15}$$

and so $\gamma_j(y\mathbf{e}_1) - \gamma_j(z\mathbf{e}_1)$ must equal

$$(y-z)\gamma_j^{(1)} + \sum_{2 \leq r \leq k - k_j} \frac{(y^{r-1} + \cdots + z^{r-1})}{r!} \sum_{\substack{t \\ k_t = k_j + r - 1}} \lambda_{tj}^{r-1,0} (y-z)\gamma_t^{(1)}. \tag{3.16}$$

Let $L_1 = L_1(\Phi)$ be a positive integer such that $L_1 \lambda_{ij}^{r,0}$ is an integer for all i, j and r. It follows almost immediately from the identity (3.16) and induction on the difference $k - k_j$, that there exists a constant $C_1 = C_1(\Phi)$ such that if (3.11) holds with $m = 1$, then for any $2 \leq j \leq N$ we have

$$\Big\| L_1 k! \gamma_j^{(1)} (y - z) \Big\| \leq C_1 X^{-k_j}. \tag{3.17}$$

Define the linear map

$$A_d : (\beta_i)_{\substack{i \in I_1 \\ k_i = d}} \mapsto \Big(\sum_{\substack{i \in I_1 \\ k_i = d}} \beta_i \lambda_{ij}^{1,0} \Big)_{\substack{2 \leq j \leq N, \\ k_j = d-1}}$$

so that

$$\Big(\gamma_j^{(1)}\Big)_{k_j = d-1} = (\alpha_i)_{\substack{i \in I_1 \\ k_i = d}} \cdot A_d.$$

We claim that each A_d is non-singular. To this end suppose that $\boldsymbol{\beta} A_d = 0$. Then a little manipulation shows that

$$\sum_{\substack{i \in I_1 \\ k_i = d}} \beta_i F_i^{1,0} = \sum_{\substack{j \\ k_j = d-1}} \Big(\sum_{\substack{i \in I_1 \\ k_i = d}} \beta_i \lambda_{ij}^{1,0} \Big) F_j = 0.$$

It follows that

$$\sum_{\substack{i \in I_1 \\ k_i = d}} \beta_i F_i(x_1, x_2) \in \mathbb{Q}[x_2].$$

This contradicts Definition 3.5, unless $\boldsymbol{\beta} = 0$. Hence for each d there exists a rational matrix B_d such that $A_d B_d = (I \mid 0)$, where I denotes the identity matrix. We therefore have that

$$\Big(\gamma_j^{(1)}\Big)_{k_j = d-1} \cdot B_d = (\alpha_i)_{\substack{i \in I_1 \\ k_i = d}} \cdot (I \mid 0). \tag{3.18}$$

Let $L_2 \in \mathbb{N}$ be such that all the matrices $L_2 B_d$ $(2 \leq k \leq k)$ have only integer entries. Then by (3.18) and (3.17), for all $i \in I_1$ we have

$$\| L_2 L_1 k! \alpha_i (y - z) \| \ll_\Phi X^{1 - k_i}.$$

Taking $L = L_2 L_1 k!$, we obtain the lemma. □

The next lemma combines Lemma 3.3 and Lemma 3.6, a combination we record since we use it repeatedly in the proof of Theorem 3.2.

Lemma 3.7. *Fix $m \in \{1,2\}$ and let C and L be as in Lemma 3.6. Suppose there exists a real $D \geq 1$ such that for any $y \in [X]$, there are at most D elements $z \in [X]$ satisfying*

$$\|L\alpha_i(y-z)\| \leq CX^{1-k_i} \quad \text{for all } i \in I_m. \tag{3.19}$$

Then we have

$$|f(\boldsymbol{\alpha};X)| \ll X^2 \log(2X)^2 \big(X^{\Delta} D/X\big)^{1/(2s)}. \tag{3.20}$$

Proof. By Lemma 3.3, it remains to prove that our assumptions imply the existence of a set $\mathcal{S} \subset [X]$ of size $|\mathcal{S}| \gg XD^{-1}$ such that for any $y, z \in \mathcal{S}$ with $y \neq z$ there exists $2 \leq j \leq N$ with

$$\|\gamma_j(y\mathbf{e}_m) - \gamma_j(z\mathbf{e}_m)\| > X^{-k_j}. \tag{3.21}$$

By Lemma 3.6, the spacing (3.21) holds for some j if there exists $i \in I_m$ such that $\|L\alpha_i(y-z)\| > CX^{1-k_i}$. Define G to be the graph on vertex set $[X]$ with y adjacent to z if and only if $\|L\alpha_i(y-z)\| \leq CX^{1-k_i}$ for all $i \in I_m$. This graph has maximal degree at most $D-1$, so (by the greedy algorithm) contains an independent set of vertices $\mathcal{S}$ of size at least $\lfloor X \rfloor / D$ (as required). □

In order to use Lemma 3.7 to relate the size of $f(\boldsymbol{\alpha})$ to the simultaneous Diophantine approximation of *all* the α_i, including those with $k_i = 1$, we must utilise major arc information. This necessitates the discussion of the standard major arc auxiliary approximation to $f(\boldsymbol{\alpha})$.

Definition 3.8. Define

$$S(q,\mathbf{a}) = \sum_{\mathbf{z}\in[q]^2} e\left(q^{-1}\mathbf{a}\cdot\mathbf{F}(\mathbf{z})\right), \quad I(\boldsymbol{\beta};X) = \int_{[0,X]^2} e\left(\boldsymbol{\beta}\cdot\mathbf{F}(\boldsymbol{\gamma})\right) \mathrm{d}\boldsymbol{\gamma}$$

and

$$V(\boldsymbol{\alpha};q,\mathbf{a}) = q^{-2}S(q,\mathbf{a})I(\boldsymbol{\alpha}-\mathbf{a}/q;X).$$

The following three results, which bound $S(q,\mathbf{a})$, $I(\boldsymbol{\beta};X)$ and the difference $f(\boldsymbol{\alpha}) - V(\boldsymbol{\alpha};q,a)$, prove useful both in this section and the next.

Lemma 3.9. *Let $q \in \mathbb{N}$ and $\mathbf{a} \in \mathbb{Z}^N$. Then for any $\varepsilon > 0$ we have*

$$S(q,\mathbf{a}) \ll_{\varepsilon} (q,\mathbf{a})^2 q^{2-\frac{1}{k}+\varepsilon}. \tag{3.22}$$

Proof. Letting $q' = (q,\mathbf{a})^{-1}q$ and $\mathbf{a}' = (q,\mathbf{a})^{-1}\mathbf{a}$, we have $S(q,\mathbf{a}) = (q,\mathbf{a})^2 S(q',\mathbf{a}')$. Hence it suffices to assume that $(q,\mathbf{a}) = 1$. Sorting the expression $\mathbf{a}\cdot\mathbf{F}$ into monomials and using the linear independence of the forms F_i, we see that there exists an integer matrix B with full row-rank such that

$$\mathbf{a}\cdot\mathbf{F}(\mathbf{x}) = \sum_{0<i_1+i_2\leq k} (\mathbf{a}B)_{\mathbf{i}}\ \mathbf{x}^{\mathbf{i}}.$$

Set $d = (\mathbf{a}B, q)$, $q' = d^{-1}q$ and $\mathbf{b}' = d^{-1}(\mathbf{a}B)$. Then $(q',\mathbf{b}') = 1$, so by [1, Lemma 8, p. 54], we have

$$\begin{aligned} S(q,\mathbf{a}) &= d^2 \sum_{\mathbf{x}\in[q']^2} e\bigg(\frac{1}{q'} \sum_{0<i_1+i_2\leq k} b'_{\mathbf{i}}\ \mathbf{x}^{\mathbf{i}}\bigg) \\ &\ll_{\varepsilon} d^2\ (q')^{2-1/k+\varepsilon} \end{aligned}$$

It therefore remains to show that $d \ll_\Phi 1$. Since B has full row-rank, there exists a rational matrix B' with

$$BB' = \Big(I \mid 0 \Big), \tag{3.23}$$

where I is the identity matrix. Clearly there exists a positive integer $m = O_\Phi(1)$ such that mB' has only integer entries. Hence we have

$$(ma_1, \dots, ma_N, 0, \dots, 0) = (\mathbf{a}B)(mB') \equiv 0 \bmod d.$$

So d divides ma_i for all i. Since $d|q$ and $(q, \mathbf{a}) = 1$, we have $d|m$. Thus $d \ll_\Phi 1$. □

Lemma 3.10. *For any $\varepsilon > 0$ the auxiliary function $I(\boldsymbol{\beta}, X)$ satisfies*

$$I(\boldsymbol{\beta}; X) \ll_\varepsilon X^2 \Big(1 + X^{k_1}|\beta_1| + \dots + X^{k_N}|\beta_N| \Big)^{-\frac{1}{k}+\varepsilon} \tag{3.24}$$

Proof. Let B be the matrix in the proof of Lemma 3.9. Changing variables in the integral $I(\boldsymbol{\beta}; X)$ gives

$$I(\boldsymbol{\beta}; X) = X^2 \int_0^1 \int_0^1 e\Big(\sum_{0 < i_1 + i_2 \leq k} X^{i_1+i_2} (\boldsymbol{\beta}B)_{\mathbf{i}}\ \boldsymbol{\gamma}^{\mathbf{i}} \Big) \mathrm{d}\boldsymbol{\gamma}. \tag{3.25}$$

Let $\beta_i(X) = X^{k_i}\beta_i$, and $\alpha_{\mathbf{i}} = (\boldsymbol{\beta}(X) \cdot B)_{\mathbf{i}} = X^{i_1+i_2}(\boldsymbol{\beta}B)_{\mathbf{i}}$. Then we can apply [1, Lemma 2, p. 50] to the double integral in (3.25) to obtain

$$I(\boldsymbol{\beta}; X) \ll_\varepsilon X^2 \min\big\{1,\ |\boldsymbol{\alpha}|_\infty^{-1/k+\varepsilon}\big\}, \tag{3.26}$$

where $|\boldsymbol{\alpha}|_\infty = \max_{\mathbf{i}} |\alpha_{\mathbf{i}}|$. Using (3.23), we have $|\boldsymbol{\beta}(X)|_\infty \ll_\Phi |\boldsymbol{\alpha}|_\infty$. The result now follows. □

Lemma 3.11. *Let q be a positive integer. Then for any $\mathbf{a} \in \mathbb{Z}^n$ and $\boldsymbol{\alpha} \in \mathbb{T}^n$*

$$f(\boldsymbol{\alpha}) - V(\boldsymbol{\alpha}; q, \mathbf{a}) \ll X \left(q + \sum_{i=1}^n |q\alpha_i - a_i| X^{k_i} \right). \tag{3.27}$$

Proof. Write $\boldsymbol{\alpha} = \mathbf{a}/q + \boldsymbol{\beta}$. Sorting the sum $f(\boldsymbol{\alpha})$ into a sum over congruence classes modulo q, we have

$$f(\boldsymbol{\alpha}) = \sum_{\mathbf{r} \in [q]^2} e\left(\mathbf{a} \cdot \mathbf{F}(\mathbf{r})/q\right) \sum_{0 \leq y_1 \leq \frac{X - r_1}{q}} \sum_{0 \leq y_2 \leq \frac{X - r_2}{q}} e\left(\boldsymbol{\beta} \cdot \mathbf{F}(q\mathbf{y} + \mathbf{r})\right). \tag{3.28}$$

By the mean value inequality we have

$$e\left(\boldsymbol{\beta} \cdot \mathbf{F}(q\mathbf{y} + \mathbf{r})\right) - q^{-2} \int_{qy_1}^{q(y_1+1)} \int_{qy_2}^{q(y_2+1)} e\left(\boldsymbol{\beta} \cdot \mathbf{F}(\boldsymbol{\gamma})\right) \mathrm{d}\boldsymbol{\gamma} \ll \sum_{i=1}^N |q\beta_i| X^{k_i - 1}.$$

Summing over $\mathbf{r}$ and $\mathbf{y}$ shows that $f(\boldsymbol{\alpha})$ equals

$$q^{-2} \sum_{\mathbf{r} \in [q]^2} e\left(\mathbf{a} \cdot \mathbf{F}(\mathbf{r})/q\right) \int_0^{X(r_1)} \int_0^{X(r_2)} e\left(\boldsymbol{\beta} \cdot \mathbf{F}(\boldsymbol{\gamma})\right) \mathrm{d}\boldsymbol{\gamma} + O\left(X \sum_{i=1}^N |q\beta_i| X^{k_i} \right), \tag{3.29}$$

where $X(r) = q\left(\left\lfloor \frac{X-r}{q} \right\rfloor + 1 \right)$. Using the fact that $|X - X(r)| \leq q$ for all $r \in [q]$, we see that (3.29) is equal to

$$q^{-2} \sum_{\mathbf{r} \in [q]^2} e\left(\mathbf{a} \cdot \mathbf{F}(\mathbf{r})/q\right) \int_0^X \int_0^X e\left(\boldsymbol{\beta} \cdot \mathbf{F}(\boldsymbol{\gamma})\right) \mathrm{d}\boldsymbol{\gamma} + O\left(Xq + X \sum_{i=1}^N |q\beta_i| X^{k_i} \right),$$

as required. □

With these bounds in hand, we are able to prove the theorem advertised at the start of this section.

Proof of the Theorem 3.2. Let $\tau = 2s\sigma + \Delta_s + \varepsilon_1$, with ε_1 sufficiently small (to be determined later). The result is vacuous if $\sigma \leq 0$, so we may assume that $\Delta_s \leq 1/3$. It then follows that $\tau < 1$. By Dirichlet's principle, for each i with $k_i \geq 2$ we can find co-prime integers b_i, q_i with $1 \leq q_i \leq X^{k_i - \tau}$ and

$$|\alpha_i - b_i/q_i| \leq q_i^{-1} X^{\tau - k_i}. \tag{3.30}$$

Let C_1 be the absolute constant in Lemma 3.7. Using (3.30), notice that if $y, z \in [X]$ satisfy (3.19), then we have

$$\|L(y-z)b_i/q_i\| \leq C_1 X^{1-k_i} + LX^{1+\tau-k_i} q_i^{-1}. \tag{3.31}$$

For each choice $y \in [X]$, the number of residue classes modulo q_i containing some $z \in [X]$ satisfying (3.31) is at most $C_1 X^{1-k_i} q_i + LX^{1+\tau-k_i} + 1$. Let D denote the maximum, over all $y \in [X]$, for the number of choices for $z \in [X]$ satisfying (3.19). Then we have

$$\begin{aligned} D &\leq (C_1 X^{1-k_i} q_i + LX^{1+\tau-k_i} + 1)(q_i^{-1} LX + 1) \\ &\ll q_i X^{1-k_i} + Xq_i^{-1} + 1. \end{aligned}$$

Using Lemma 3.7, we see that

$$|f(\boldsymbol{\alpha})| \ll X^2 \log(2X)^2 \left(X^{\Delta} (q_i X^{-k_i} + q_i^{-1} + X^{-1}) \right)^{1/(2s)}.$$

By the lower bound (3.1), we have

$$\begin{aligned} X^{-2s\sigma - \Delta_s} &\ll (q_i X^{-k_i} + q_i^{-1} + X^{-1}) \log(2X)^{4s} \\ &\ll (X^{-\tau} + q_i^{-1}) \log(2X)^{4s}. \end{aligned}$$

Since $\tau > 2s\sigma + \Delta_s$, we must have

$$q_i \ll X^{2s\sigma + \Delta_s} \log(2X)^{4s}. \tag{3.32}$$

Since $1 > 2s\sigma + \Delta_s$, this implies that the right-hand side of (3.31) is strictly less than q_i^{-1} (provided $X \gg_\Phi 1$). It follows that (3.31) implies $q_i | L(y-z)b_i$, which in turn implies $q_i | L(y-z)$, since $(q_i, b_i) = 1$. Hence it follows from the assumption (3.19) that $q_i | L(y-z)$ for all $i \in I_m$. Let Q_m denote the lowest common multiple of the set $\{q_i : i \in I_m\}$. Then the number D satisfies $D \ll XQ_m^{-1} + 1$. Utilising Lemma 3.7 again, we obtain the bound

$$Q_m \ll X^{2s\sigma + \Delta_s} \log(2X)^{4s}. \tag{3.33}$$

Let $Q = [Q_1, Q_2]$, so that $Q \ll X^{4s\sigma + 2\Delta_s} \log(2X)^{8s}$ by (3.33). Since σ is strictly less than $\frac{1-3\Delta_s}{6s+3}$, we can (on taking ϵ_1 sufficiently small) find a real μ satisfying

$$\sigma < \mu < \tfrac{1}{2} \left(1 - (4s+1)\sigma - 2\Delta - \tau \right). \tag{3.34}$$

As the space of linear binary homogeneous polynomials has dimension 2, there are at most two indices i with $k_i = 1$. We can therefore use Dirichlet's principle to find a positive integer $1 \leq t \leq X^{2\mu}$, along with $a_i \in \mathbb{Z}$ $(k_i = 1)$ which are together co-prime to t and such that

$$|t(Q\alpha_i) - a_i| \leq X^{-\mu} \quad (k_i = 1). \tag{3.35}$$

Set $q = tQ$. For i with $k_i \geq 2$, let us define $a_i = (q/q_i)b_i$. Then the N-tuple $\mathbf{a} = (a_1, \dots, a_N)$ satisfies $(q, \mathbf{a}) = 1$ and

$$|q\alpha_i - a_i| \ll \begin{cases} X^{-\mu} & (k_i = 1), \\ X^{2\mu+4s\sigma+2\Delta_s+\tau-k_i} \log(2X)^{4s} & (k_i \geq 2). \end{cases} \tag{3.36}$$

It thus follows from Lemma 3.11 and (3.34) that

$$\begin{aligned} |f(\boldsymbol{\alpha}) - V(\boldsymbol{\alpha}; q, \mathbf{a})| &\ll X^{2-\mu} + X^{1+2\mu+4s\sigma+2\Delta_s+\tau} \log(2X)^{4s} \\ &= o\Big(X^{2-\sigma}\Big). \end{aligned}$$

Hence by the lower bound (3.1), we have $|V(\boldsymbol{\alpha}; q, \mathbf{a})| \gg X^{2-\sigma}$. Combining this, together with Lemma 3.9 and Lemma 3.10 , we see that for any $\varepsilon > 0$ we have

$$q + \sum_{i=1}^{N} |q\alpha_i - a_i| X^{k_i} \ll X^{k\sigma + \frac{\varepsilon}{2}}.$$

Taking X sufficiently large (in terms of s, ε and Φ), we obtain the theorem. □

4. The Asymptotic Formula

In order to prove our density result, Theorem 1.3, we need to estimate the number of solutions to (1.5) when the variables $\mathbf{x}_j$ are restricted to the box $[X]^2$.

Definition 4.1. Given a finite set $A \subset \mathbb{Z}^2$, write $R_{\mathbf{c},\Phi}(A)$ for the number of tuples $(\mathbf{x}_1, \dots, \mathbf{x}_s)$ in the set A^s satisfying

$$c_1 \Phi^{u,v}(\mathbf{x}_1) + \cdots + c_s \Phi^{u,v}(\mathbf{x}_s) = 0 \quad (0 \leq u + v < k). \tag{4.1}$$

When $A = [X]^2$, we simply write $R_{\mathbf{c},\Phi}(X)$.

The Hardy–Littlewood method gives an asymptotic for $R_{\mathbf{c},\Phi}(X)$, an asymptotic whose main term is a product of local densities, which we now define.

Definition 4.2. Let Φ denote a binary form of degree k, differential dimension N and differential degree K. Let $\{F_1, \dots, F_N\}$ denote a maximal linearly independent subset of $\{\Phi^{u,v} : 0 \leq u + v < k\}$. When $T > 0$, define $\lambda_T(y) = T \max\{0, 1 - T|y|\}$ and

$$\mu_T = \mu_T(\mathbf{c}) = \int_{[0,1]^{2s}} \lambda_T\Big(\sum_{j=1}^{s} c_j F_1(\boldsymbol{\gamma}_j)\Big) \cdots \lambda_T\Big(\sum_{j=1}^{s} c_j F_N(\boldsymbol{\gamma}_j)\Big) \mathrm{d}\boldsymbol{\gamma}$$

The limit $\sigma_\infty = \sigma_\infty(\mathbf{c}) = \lim_{T\to\infty} \mu_T$, when it exists, is called the *real density.* Given a natural number q, we write

$$M(q) = M(q; \mathbf{c}) = \#\Big\{\mathbf{x} \in (\mathbb{Z}/q\mathbb{Z})^{2s} : \sum_{j=1}^{s} c_j \mathbf{F}(\mathbf{x}_j) \equiv 0 \pmod{q}\Big\}.$$

For each prime p, the limit

$$\sigma_p(\mathbf{c}) = \lim_{H\to\infty} p^{-H(2s-N)} M(p^H), \tag{4.2}$$

provided it exists, is called the *p-adic density.*

The purpose of this section is to prove the following asymptotic formula.

Theorem 4.3. *Let Φ be a non-degenerate binary form of degree k, differential dimension N and differential degree K. Suppose that*

$$s \geq kN(\log K + \log\log K + 26). \tag{4.3}$$

Then there exists $\delta > 0$ such that for any choice of non-zero integers $c_1, \dots, c_s$ we have

$$R_{\mathbf{c},\Phi}(X) = \sigma_\infty\Big(\prod_p \sigma_p\Big)X^{2s-K} + O(X^{2s-K-\delta}) \tag{4.4}$$

Suppose in addition that $\mathbf{c}$ is a non-singular choice of coefficients for Φ. Then

$$\sigma_\infty \prod_p \sigma_p > 0. \tag{4.5}$$

Remark 4.4. The $O(1)$ constant in (4.3) can certainly be lowered from 26 if one is willing to implement the results of §3 more optimally.

The proof of Theorem 4.3 proceeds by the usual Hardy–Littlewood dissection into major and minor arcs.

Definition 4.5. Given a tuple of integers $\mathbf{a} = (a_1, \dots, a_N)$ and $q \in \mathbb{N}$, define the *major arc* centred at $\mathbf{a}/q$ to be the set

$$\mathfrak{M}(q, \mathbf{a}) = \Big\{\boldsymbol{\alpha} \in \mathbb{T}^N : \|\alpha_i - a_i/q\| \leq q^{-1}X^{\frac{1}{4}-k_i} \quad (1 \leq i \leq N)\Big\}.$$

Define the *major arcs* $\mathfrak{M}$ to be the union of the sets $\mathfrak{M}(q, \mathbf{a})$ with $1 \leq q \leq X^{1/4}$ and $\mathbf{a} \in [q]^n$ subject to $(q, \mathbf{a}) = 1$. Define the *minor arcs* to be the complement $\mathfrak{m} = \mathbb{T}^N \setminus \mathfrak{M}$.

One can show that for $X \gg 1$ the major arcs are disjoint. We can therefore define the function $V(\boldsymbol{\alpha})$ to equal $V(\boldsymbol{\alpha}; q, \mathbf{a})$ when $\boldsymbol{\alpha} \in \mathfrak{M}(q, \mathbf{a}) \subset \mathfrak{M}$, and equal 0 otherwise.

Lemma 4.6. *Whenever $s \geq k(N+1)+1$ there exists $\delta > 0$ such that*

$$\int_{\mathfrak{M}} V(c_1\boldsymbol{\alpha}) \cdots V(c_s\boldsymbol{\alpha}) \mathrm{d}\boldsymbol{\alpha} = \mathfrak{J}\mathfrak{S}X^{2s-K} + O(X^{2s-K-\delta}), \tag{4.6}$$

where

$$\mathfrak{J} = \mathfrak{J}(\mathbf{c}) = \int_{\mathbb{R}^N}\int_{[0,1]^{2s}} e\Big(\boldsymbol{\beta} \cdot \sum_{j=1}^{s} c_j \mathbf{F}(\boldsymbol{\gamma}_j)\Big)\mathrm{d}\boldsymbol{\gamma}\mathrm{d}\boldsymbol{\beta} \tag{4.7}$$

and

$$\mathfrak{S} = \mathfrak{S}(\mathbf{c}) = \sum_{q=1}^{\infty} q^{-2s} \sum_{\substack{\mathbf{a}\in[q]^N \\ (q,\mathbf{a})=1}} S(q, c_1\mathbf{a}) \cdots S(q, c_s\mathbf{a}). \tag{4.8}$$

Proof. Define $A(q)$ to be the sum

$$A(q) = \sum_{\substack{\mathbf{a}\in[q]^N \\ (q,\mathbf{a})=1}} q^{-2s} S(q, c_1\mathbf{a}) \cdots S(q, c_s\mathbf{a}), \tag{4.9}$$

and let $\mathcal{I}(\boldsymbol{\beta}; X)$ denote the product

$$\mathcal{I}(\boldsymbol{\beta}; X) = I(c_1\boldsymbol{\beta}; X) \cdots I(c_s\boldsymbol{\beta}; X).$$

Then by disjointness of the major arcs, and a change of variables $\beta_i = (\alpha_i - a_i/q)X^{k_i}$, we have

$$\int_{\mathfrak{M}} V(c_1\boldsymbol{\alpha})\cdots V(c_s\boldsymbol{\alpha})\mathrm{d}\boldsymbol{\alpha} = X^{-K} \sum_{1\le q\le X^{1/4}} A(q) \int_{\mathcal{B}_q} \mathcal{I}(\beta_1 X^{-k_1}, \dots, \beta_N X^{-k_N}; X)\mathrm{d}\boldsymbol{\beta}, \tag{4.10}$$

where $\mathcal{B}_q = \prod_{1\le i\le N}[-q^{-1}X^{1/4}, q^{-1}X^{1/4}]$. Let $\delta_1 = \frac{1}{5}(\frac{s}{kN} - 1) > 0$. By Lemma 3.10 and the AM–GM inequality we have

$$\mathcal{I}(\beta_1 X^{-k_1}, \dots, \beta_N X^{-k_N}; X) \ll X^{2s} \prod_{i=1}^{N} \big(1 + |\beta_i|\big)^{-\frac{s}{kN}+\delta_1}. \tag{4.11}$$

It follows that

$$\begin{aligned}\int_{\mathbb{R}^N\setminus\mathcal{B}_q} &\mathcal{I}(\beta_1 X^{-k_1}, \dots, \beta_N X^{-k_N}; X)\mathrm{d}\boldsymbol{\beta} \\ &\ll X^{2s} \int_{X^{1/4}}^{\infty} x^{-(1+4\delta_1)}\mathrm{d}x \left(\int_{\mathbb{R}} \frac{1}{(1+|x|)^{1+4\delta_1}}\mathrm{d}x\right)^{s-1} \\ &\ll_{s,\delta_1} X^{2s-\delta_1}.\end{aligned}$$

Combining this with another change of variables, we have

$$\int_{\mathcal{B}_q} \mathcal{I}(\beta_1 X^{-k_1}, \dots, \beta_N X^{-k_N}; X)\mathrm{d}\boldsymbol{\beta} = \mathfrak{J}X^{2s} + O(X^{2s-\delta_1}). \tag{4.12}$$

Set $\delta_2 = \frac{1}{5}(\frac{s}{k} - N - 1) > 0$. By Lemma 3.9

$$A(q) \ll q^{N-\frac{s}{k}+\delta_2} = q^{-1-4\delta_2}.$$

Thus

$$\sum_{q>X^{1/4}} |A(q)| \ll \sum_{q>X^{1/4}} q^{-1-4\delta_2} \ll_{\delta_2} X^{-\delta_2}.$$

It follows that

$$\sum_{1\le q\le X^{1/4}} A(q) = \mathfrak{S} + O(X^{-\delta_2}). \tag{4.13}$$

Combining (4.10), (4.12) and (4.13), we obtain the result. □

Lemma 4.7. *There exist positive integers r and t such that for any $s \ge r + 2Mt$ there exists $\tau > 0$ such that*

$$\int_{\mathfrak{m}} |f(\boldsymbol{\alpha})|^s \mathrm{d}\boldsymbol{\alpha} \ll X^{2s-K-\tau}. \tag{4.14}$$

Moreover, one can ensure that

$$r + 2tM \le kN(\log K + \log\log K + 26).$$

Proof. Let us first find a large value for the expression $\frac{1-3\Delta_s}{6s+3}$ occurring in Theorem 3.2. Setting $s_0 = M\lceil k\log(21K)\rceil$, by Theorem 1.7 we have

$$\Delta_{s_0} < Ke^{-\lceil k\log(21K)\rceil/k} \le \tfrac{1}{21}.$$

Therefore

$$\tfrac{1-3\Delta_{s_0}}{6s_0+3} > \tfrac{1}{7s_0+(7/2)} := \sigma, \text{ say.}$$

Since Φ is non-degenerate of degree at least two, Lemma 2.2 guarantees that Φ has two linearly independent derivatives of degree one. This implies that $N \geq 3$ and $K \geq 4$. Hence

$$k\sigma = \tfrac{k}{7s_0+(7/2)} \leq \tfrac{1}{M\log(21K)} \leq \tfrac{1}{2.\log(21.4)} < \tfrac{1}{8}. \tag{4.15}$$

Let $\boldsymbol{\alpha} \in \mathfrak{m}$ and suppose that

$$|f(c_j\boldsymbol{\alpha})| \geq X^{2-\sigma}. \tag{4.16}$$

Provided X is sufficiently large, it follows from Theorem 3.2 and (4.15) that there exists $q \in \mathbb{N}$ and integers $a_1, \ldots, a_N$, with $|q(c_j\alpha_i) - a_i| \leq X^{1/8}$ and $q \leq X^{1/8}$. For $X \gg_{\mathbf{c}} 1$ sufficiently large, we have $|c_j|q \leq |c_j|X^{1/8} \leq X^{1/4}$, so $\boldsymbol{\alpha} \in \mathfrak{M}(c_jq, \mathbf{b}) \subset \mathfrak{M}$, a contradiction. Hence we must in fact have

$$|f(c_j\boldsymbol{\alpha})| \leq X^{2-\sigma}. \tag{4.17}$$

Set

$$t = \lceil k\log(K\log K)\rceil \quad \text{and} \quad r = \left\lceil \sigma^{-1}Ke^{-t/k}\right\rceil, \tag{4.18}$$

and let Δ_{tM} be an admissible exponent for (tM, Φ). It suffices to prove (4.14) for $s = r + 2tM$. By (4.17), Hölder's inequality and Theorem 1.7, we have

$$\begin{aligned}\int_{\mathfrak{m}} f(c_1\boldsymbol{\alpha})\cdots f(c_s\boldsymbol{\alpha})\mathrm{d}\boldsymbol{\alpha} &\leq X^{2r-r\sigma}\oint |f(\boldsymbol{\alpha})|^{2tM}\mathrm{d}\boldsymbol{\alpha}\\ &\ll X^{2s-K-(r\sigma-\Delta_{tM})}.\end{aligned}$$

Since $r\sigma > \Delta_{tM}$, we obtain (4.14).

It remains to show that $r + 2tM \leq kN(\log K + \log\log K + 21)$. Using the fact that $k \geq 2$, $N \geq 3$, $K \geq 4$ and $K \leq \min\{N^2, k^3\}$, we have

$$\begin{aligned}2tM &\leq t(N+1)\\ &\leq k(N+1)\log(K\log K) + N + 1\\ &\leq kN(\log K + \log\log K + 4)\end{aligned}$$

and

$$\begin{aligned}r &\leq Ke^{-t/k}\sigma^{-1} + 1\\ &\leq \frac{7kM\log(21K) + 7M + (7/2)}{\log K} + 1\\ &\leq 22kN.\end{aligned}$$

□

Proof of Theorem 4.3. Let r and t be defined as in the proof of Lemma 4.7. We deduce the theorem under the weaker assumption that $s \geq r + 2tM$. From this assumption, it follows that $s = u + 2vM$, where $u \geq 1$ and $v \geq kM(K\log(K\log K) + 6)$. Therefore $\Delta_{vM} \leq e^{-6} < 3/4$. Combining this with Theorem 1.7, Lemma 3.11

and Hölder's inequality, we see that there exists $j \in [s]$ such that

$$\begin{aligned}\int_{\mathfrak{M}} (f(c_1\boldsymbol{\alpha})\cdots f(c_s\boldsymbol{\alpha}) - V(c_1\boldsymbol{\alpha})\cdots V(c_s\boldsymbol{\alpha}))\mathrm{d}\boldsymbol{\alpha}&\\ &\ll X^{2u-\frac{3}{4}}\Big(\oint |f(\boldsymbol{\alpha})|^{2vM}\mathrm{d}\boldsymbol{\alpha} + \int_{\mathfrak{M}} |V(c_j\boldsymbol{\alpha})|^{2vM}\mathrm{d}\boldsymbol{\alpha}\Big)\\ &\ll X^{2s-K+\Delta_{vM}-\frac{3}{4}} + X^{2s-K-\frac{3}{4}}\\ &\ll X^{2s-K-\tau_1}, \quad \text{say.}\end{aligned}$$

Using this, together with Lemma 4.6 and Lemma 4.7, we see there exists $\tau_2 > 0$ such that

$$\oint f(c_1\boldsymbol{\alpha})\cdots f(c_s\boldsymbol{\alpha})\mathrm{d}\boldsymbol{\alpha} = \mathfrak{S}\mathfrak{J}X^{2s-K} + O(X^{2s-K-\tau_2}) \tag{4.19}$$

It remains to show that $\mathfrak{J} = \sigma_\infty$, that $\mathfrak{S} = \prod_p \sigma_p$ and that these quantities are positive under the appropriate non-singularity hypotheses. To prove $\mathfrak{J} = \sigma_\infty$ we use a method of Schmidt [16], as described by Parsell [12]. For a positive real T, define

$$K_T(\beta) = \left(\frac{\sin(\pi\beta T^{-1})}{\pi\beta T^{-1}}\right)^2, \qquad \mathcal{K}_T(\boldsymbol{\beta}) = K_T(\beta_1)\cdots K_T(\beta_n).$$

It follows from Baker [2, Lemma 14.1] that

$$\begin{aligned}\hat{K}_T(y) &= \int_{\mathbb{R}} K_T(\beta)e(-\beta y)d\beta\\ &= T\max\{0, 1 - T|y|\}.\end{aligned} \tag{4.20}$$

In particular, this Fourier transform is always non-negative. Write

$$I(\boldsymbol{\beta}) = I(\boldsymbol{\beta}; 1) = \int_{[0,1]^2} e(\boldsymbol{\beta}\cdot\mathbf{F}(\gamma))\, d\boldsymbol{\gamma} \qquad \text{and} \qquad \mathcal{I}(\boldsymbol{\beta}) = I(c_1\boldsymbol{\beta})\cdots I(c_s\boldsymbol{\beta}).$$

By Fubini's theorem, we have that

$$\begin{aligned}\mu_T &= \int_{[0,1]^{2s}} \hat{K}_T\Big(\sum_{i=1}^{s} c_iF_1(\boldsymbol{\gamma}_i)\Big)\cdots\hat{K}_T\Big(\sum_{i=1}^{s} c_iF_N(\boldsymbol{\gamma}_i)\Big)\mathrm{d}\boldsymbol{\gamma}\\ &= \int_{\mathbb{R}^N} \mathcal{K}_T(\boldsymbol{\beta})\mathcal{I}(\boldsymbol{\beta})\mathrm{d}\boldsymbol{\beta}\end{aligned}$$

Lemma 3.10 and the AM–GM inequality ensure that, for any $\varepsilon > 0$, we have the bound

$$\mathcal{I}(\boldsymbol{\beta}) \ll_\varepsilon \prod_{1\le i\le N}(1+|\beta_i|)^{-\frac{s}{kN}+\varepsilon},$$

and a simple estimate reveals that

$$1 - \mathcal{K}_T(\boldsymbol{\beta}) \ll \min\{1, |\boldsymbol{\beta}|^2T^{-2}\}. \tag{4.21}$$

Therefore

$$\begin{aligned}\mathfrak{J} - \mu_T &\ll \int_{\mathbb{R}^N} \min\{1, |\boldsymbol{\beta}|^2T^{-2}\}\prod_{1\le i\le N}(1+|\beta_i|)^{-\frac{s}{kN}+\varepsilon}\mathrm{d}\boldsymbol{\beta}\\ &\ll \int_{|\boldsymbol{\beta}|>T^{\frac{1}{3N}}}\prod_{1\le i\le N}(1+|\beta_i|)^{-\frac{s}{kN}+\varepsilon}\mathrm{d}\boldsymbol{\beta} + \int_{|\boldsymbol{\beta}|\le T^{\frac{1}{3N}}}|\boldsymbol{\beta}|^2T^{-2}\mathrm{d}\boldsymbol{\beta}.\end{aligned}$$

Using $s > kN$, we see that

$$\mathfrak{J} = \lim_{T\to\infty}\mu_T = \sigma_\infty.$$

Next, let us suppose that there exists a non-singular real solution to (1.5). Writing $\mathbf{P}(\mathbf{x})$ for $\sum_{i=1}^{s} c_i \mathbf{F}(x_{2i-1}, x_{2i})$, it follows that there is some $\boldsymbol{\xi} \in \mathbb{R}^{2s}$ for which $\mathbf{P}(\boldsymbol{\xi}) = 0$, along with $S \subset [2s]$ such that $|S| = N$ and

$$\det \left(\frac{\partial P_i}{\partial \xi_j}(\boldsymbol{\xi}) \right)_{\substack{1 \le i \le N \\ j \in S}} \neq 0.$$

The translation-dilation invariance of (1.5) ensures that we may assume that $\boldsymbol{\xi} \in (0,1)^{2s}$. Let $[2s] \setminus S = \{l(N+1), \ldots, l(2s)\}$. Define the function $\rho : \mathbb{R}^{2s} \to \mathbb{R}^{2s}$ by

$$\rho_i(\boldsymbol{\gamma}) = \begin{cases} P_i(\boldsymbol{\gamma}) & \text{if } 1 \le i \le N, \\ \gamma_{l(i)} & \text{if } N < i \le 2s. \end{cases}$$

Let $\boldsymbol{\eta} = \rho(\boldsymbol{\xi})$, so that $\eta_i = 0$ for $1 \le i \le N$. Notice that

$$|\det \rho'(\boldsymbol{\gamma})| = \left| \det \left(\frac{\partial P_i}{\partial \xi_j}(\boldsymbol{\gamma}) \right)_{1 \le i \le N,\, j \in S} \right|.$$

By the Inverse Function Theorem, there exists an open set $U \subset [0,1]^{2s}$ which contains $\boldsymbol{\xi}$ and an open set V containing $\boldsymbol{\eta}$ such that ρ is a homeomorphism from U to V. Define the constant $C_1 = C_1(\Phi, s)$ by

$$C_1 = \max_{\boldsymbol{\gamma} \in [0,1]^{2s}} \left| \det \left(\frac{\partial P_i}{\partial \xi_j}(\boldsymbol{\gamma}) \right)_{1 \le i \le N,\, j \in S} \right|.$$

Using positivity of the Fourier transform $\hat{K}_T$ and the fact that $U \subset [0,1]^{2s}$, we have

$$\int_{[0,1]^{2s}} \hat{K}_T(P_1(\boldsymbol{\gamma})) \cdots \hat{K}_T(P_N(\boldsymbol{\gamma})) \mathrm{d}\boldsymbol{\gamma} \ge \int_U \hat{K}_T(P_1(\boldsymbol{\gamma})) \cdots \hat{K}_T(P_N(\boldsymbol{\gamma})) \mathrm{d}\boldsymbol{\gamma}.$$

This latter integral is in turn bounded below by

$$\frac{1}{C_1} \int_U \hat{K}_T(\rho_1(\boldsymbol{\gamma})) \cdots \hat{K}_T(\rho_N(\boldsymbol{\gamma})) |\det \rho'(\boldsymbol{\gamma})| \mathrm{d}\boldsymbol{\gamma}.$$

By a change of variables this equals

$$\frac{1}{C_1} \int_V \hat{K}_T(\zeta_1) \cdots \hat{K}_T(\zeta_N) \mathrm{d}\boldsymbol{\zeta}.$$

Since V is defined independently of T, there exists $\varepsilon = \varepsilon(\Phi, s) > 0$ such that if $|\zeta_i - \eta_i| \le \varepsilon$ $(1 \le i \le 2s)$, then $\boldsymbol{\zeta} \in V$. Let W_T denote the set of $\boldsymbol{\zeta}$ for which $|\zeta_i| \le (2T)^{-1}$ $(1 \le i \le N)$ and $|\zeta_i - \eta_i| \le \varepsilon$ $(i > N)$. Then for $T \ge (2\varepsilon)^{-1}$, the set W_T is contained in V. Moreover, for $\boldsymbol{\zeta} \in W_T$ and $1 \le i \le N$ we have $\hat{K}_T(\zeta_i) \ge T/2$. Therefore

$$\begin{aligned} \int_V \hat{K}_T(\zeta_1) \cdots \hat{K}_T(\zeta_N) \mathrm{d}\boldsymbol{\zeta} &\ge \int_{W_T} \hat{K}_T(\zeta_1) \cdots \hat{K}_T(\zeta_N) \mathrm{d}\boldsymbol{\zeta} \\ &\ge \mathrm{meas}(W_T) \frac{T^N}{2^N} \\ &\ge T^{-N} (2\varepsilon)^{2s-N} \frac{T^N}{2^N} \\ &\gg_{\Phi,s} 1. \end{aligned}$$

Hence $\mu_T \gg_{\Phi,s} 1$ for all sufficiently large T.

Let us now turn to the singular series $\mathfrak{S}$. Recalling (4.9), for each prime p define

$$T(p) = \sum_{h=0}^{\infty} A(p^h),$$

By Lemma 3.9 this series is absolutely convergent for $s > k(N+1)$.

Let q and r be coprime positive integers. By Euclid's algorithm, any pair $\mathbf{x}$ of residues modulo qr can be represented uniquely in the form $r\mathbf{y} + q\mathbf{z}$ with $\mathbf{y} \in [q]^2$ and $\mathbf{z} \in [r]^2$. It follows that for $\mathbf{a}, \mathbf{b} \in \mathbb{Z}^N$ we have $S(qr, r\mathbf{a} + q\mathbf{b}) = S(q, \mathbf{a})S(r, \mathbf{b})$. Again, by Euclid's algorithm, each N-tuple $\mathbf{a}'$ of residues modulo qr with $(\mathbf{a}', qr) = 1$ can be represented uniquely in the form $r\mathbf{a} + q\mathbf{b}$ with $\mathbf{a} \in [q]^N$, $(\mathbf{a}, q) = 1$ and $\mathbf{b} \in [r]^N$, $(\mathbf{b}, r) = 1$. A similar argument therefore gives $A(qr) = A(q)A(r)$.

Let $p_1, \dots, p_m$ denote the primes bounded above by X. Using multiplicativity of $A(q)$, together with Lemma 3.9 and the fact that $\frac{s}{k} - N = 1 + 2\varepsilon$ for some $\varepsilon > 0$, we have

$$\begin{aligned}\prod_{p \le X} T(p) - \mathfrak{S} &= \sum_{h_1=0}^{\infty} \cdots \sum_{h_m=0}^{\infty} A(p_1^{h_1} \cdots p_m^{h_m}) - \sum_{q=1}^{\infty} A(q) \\ &\ll \sum_{q > X} |A(q)| \\ &\ll \sum_{q > X} q^{N + \varepsilon - s/k} \to 0 \qquad \text{as } X \to \infty.\end{aligned}$$

By orthogonality

$$\begin{aligned}M(p^H) &= \sum_{\mathbf{x} \in [p^H]^{2s}} p^{-NH} \sum_{\mathbf{a} \in [p^H]^N} e\left(\mathbf{a} \cdot \mathbf{P}(\mathbf{x})/p^H\right) \\ &= p^{-HN} \sum_{\mathbf{a} \in [p^H]^N} \prod_{j=1}^{s} S(p^H, c_j \mathbf{a}).\end{aligned}$$

Partitioning the sum over $\mathbf{a}$ according to the value of $(p^H, \mathbf{a})$, we see that $M(p^H)$ is equal to

$$p^{H(2s-N)} \sum_{h=0}^{H} A(p^h).$$

It follows that $\mathfrak{S} = \prod_p \sigma_p$.

To show positivity of $\mathfrak{S}$, we begin with the following result from elementary linear algebra.

Lemma 4.8. *Let h, H be non-negative integers with $H \ge h + 1$. Suppose that A is an $n \times n$ integer matrix with $p^h \| \det A$. Then the image $\left\{A \cdot \mathbf{x} : \mathbf{x} \in (\mathbb{Z}/p^H\mathbb{Z})^n\right\}$ contains the subgroup $\left\{p^h \mathbf{y} : \mathbf{y} \in (\mathbb{Z}/p^H\mathbb{Z})^n\right\}$.*

For a proof of this lemma, let A_{ij} denote the ij-minor of A, obtained from A by deleting the ith row and jth column. We define the adjunct matrix of A by

$$\mathrm{adj}(A) = \left((-1)^{i+j} A_{ji}\right)_{1 \le i,j \le n}.$$

Then we have the identity

$$A \cdot \mathrm{adj}(A) = \det(A) I_n. \tag{4.22}$$

Since $p^h || \det(A)$, we have $\det(A) = up^h$ where u is a unit in $\mathbb{Z}/p^H\mathbb{Z}$. Let $\mathbf{y} \in (\mathbb{Z}/p^H\mathbb{Z})^n$. Then

$$\begin{aligned} p^h\mathbf{y} &= \det(A)(u^{-1}\mathbf{y}) \\ &= A \cdot (u^{-1}\mathrm{adj}(A) \cdot \mathbf{y}), \end{aligned}$$

as required. This completes the proof of Lemma 4.8.

Given a subset $S \subset [2s]$, define the Jacobian matrix

$$J_{\mathbf{P}}(\mathbf{x}; S) = \left(\frac{\partial P_i}{\partial x_j}(\mathbf{x}) \right)_{1 \le i \le N,\, j \in S}.$$

When $|S| = N$ we define $\Delta_{\mathbf{P}}(\mathbf{x}; S)$ to be the determinant of $J_{\mathbf{P}}(\mathbf{x}; S)$. Given a positive integer h, let $\mathcal{B}_h(p^H)$ denote the set of $\mathbf{x} \in (\mathbb{Z}/p^H\mathbb{Z})^{2s}$ with $\mathbf{P}(\mathbf{x}) \equiv 0 \bmod p^H$ and for which there exists $S \subset [2s]$ with $|S| = N$ and $p^h || \Delta_{\mathbf{P}}(\mathbf{x}; S)$. The following claim is a version of Hensel's lemma.

Claim. *For $H \ge 2h + 1$ we have the bound*

$$\left|\mathcal{B}_h(p^{H+1})\right| \ge p^{(2s-N)} \left|\mathcal{B}_h(p^H)\right|. \tag{4.23}$$

Fix $\mathbf{x} \in (\mathbb{Z}/p^H\mathbb{Z})^{2s}$ with $\mathbf{P}(\mathbf{x}) \equiv 0 \bmod p^H$ and $S \subset [2s]$ with $|S| = N$ and $p^h || \Delta_{\mathbf{P}}(\mathbf{x}; S)$. For each $j \notin S$ choose $y_j \in [p]$ and define

$$z_j = \begin{cases} x_j & (j \in S), \\ x_j + p^H y_j & (j \notin S). \end{cases}$$

Let $\mathbf{w} \in \mathbb{Z}^{2s}$ be subject to the condition that $w_j = 0$ if $j \notin S$. By the binomial theorem and the fact that $2(H - h) \ge H + 1$, we have

$$\mathbf{P}(\mathbf{z} + p^{H-h}\mathbf{w}) \equiv \mathbf{P}(\mathbf{z}) + p^{H-h} J_{\mathbf{P}}(\mathbf{z}; S) \cdot (w_j)_{j \in S} \pmod{p^{H+1}}. \tag{4.24}$$

Since $\mathbf{P}(\mathbf{z}) \equiv \mathbf{P}(\mathbf{x}) \equiv 0 \mod p^H$, we see that p^h divides every entry in the N-tuple of integers $\mathbf{P}(\mathbf{z})/p^{H-h}$. Hence

$$-\mathbf{P}(\mathbf{z})/p^{H-h} \in \left\{ p^h\mathbf{y} : \mathbf{y} \in (\mathbb{Z}/p^{h+1}\mathbb{Z})^N \right\}.$$

Notice that $\Delta_{\mathbf{P}}(\mathbf{z}; S) \equiv \Delta_{\mathbf{P}}(\mathbf{x}; S) \mod p^{h+1}$, and so $p^h || \Delta_{\mathbf{P}}(\mathbf{z}; S)$. Therefore, by Lemma 4.8, for each $j \in S$ we can find $w_j \in \mathbb{Z}/p^{h+1}\mathbb{Z}$ so that

$$J_{\mathbf{P}}(\mathbf{z}; S) \cdot (w_j)_{j \in S} \equiv -\mathbf{P}(\mathbf{z})/p^{H-h} \pmod{p^{h+1}}.$$

Moreover, since $H - h \ge h + 1$, we have $\Delta_{\mathbf{P}}(\mathbf{z} + p^{H-h}\mathbf{w}; S) \equiv \Delta_{\mathbf{P}}(\mathbf{x}; S) \mod p^{h+1}$. Hence

$$\mathbf{z} + p^{H-h}\mathbf{w} \in \mathcal{B}_h(p^{H+1}).$$

Since $w_j = 0$ if $j \notin S$, we see that for each choice of $\mathbf{z}$, the sum $\mathbf{z} + p^{H-h}\mathbf{w}$ gives a unique element of $\mathcal{B}_h(p^{H+1})$. As there are p^{2s-N} choices for $\mathbf{z}$ for each choice of $\mathbf{x} \in \mathcal{B}_h(p^H)$, the claim follows.

Suppose there exists $\mathbf{x} \in \mathbb{Q}_p^{2s}$ such that $\mathbf{P}(\mathbf{x}) = 0$ and $S \subset [2s]$ such that the Jacobian matrix $J_{\mathbf{P}}(\mathbf{x}; S)$ is non-singular over $\mathbb{Q}_p$. By homogeneity of the P_i we may assume all the entries of $\mathbf{x}$ are p-adic integers. Hence there exists a non-negative integer h such that $|\Delta_{\mathbf{P}}(\mathbf{x}; S)|_p = p^{-h}$. Take any $\mathbf{y} \in \mathbb{Z}^t$ such that $\mathbf{y} \equiv \mathbf{x} \mod p^{2h+1}$. Then $p^h || \Delta_{\mathbf{P}}(\mathbf{y}; S)$ and $\mathbf{P}(\mathbf{y}) \equiv 0 \mod p^{2h+1}$, so $\mathbf{y} \in \mathcal{B}_h(p^{2h+1})$. In particular, $|\mathcal{B}_h(p^{2h+1})| \ge 1$. Iterating the bound (4.23) obtained in the previous lemma, we have established that there exists a non-negative integer $h = h(\Phi, p)$ such that for all $H \geq 2h + 1$ we have the lower bound

$$|\mathcal{B}_h(p^H)| \geq p^{(2s-N)(H-2h-1)}. \qquad (4.25)$$

Clearly $M(p^H) \geq \mathcal{B}_h(p^H)$, so inputting this into the relation (4.2), we obtain

$$\begin{aligned} T(p) &= \lim_{H\to\infty} p^{-H(2s-N)} M(p^H) \\ &\geq p^{-(2s-N)(2h+1)} \\ &\gg_{s,\Phi,p} 1. \end{aligned}$$

The absolute convergence of the product $\mathfrak{S} = \prod_p T(p)$, with all $T(p)$ positive, implies that

$$\lim_{X\to\infty} \prod_{p>X} T(p) = 1.$$

In particular, there exists X_0 such that for all $p > X_0$ we have

$$\prod_{p>X_0} T(p) > 1/2.$$

Since $T(p) > 0$ for all $p \leq X_0$, we also have $\prod_{p\leq X_0} T(p) > 0$. Therefore

$$\mathfrak{S} = \prod_{p\leq X_0} T(p) \prod_{p>X_0} T(p) > 0.$$

□

5. Density bounds for solution-free sets

This section is dedicated to the proof of our main theorem.

Theorem 5.1. *Let $\Phi \in \mathbb{Z}[x, y]$ be a binary form of degree $k \geq 2$, differential dimension N and differential degree K, and let $\mathbf{c} \in \mathbb{Z}^s$ be a non-singular choice of coefficients for Φ with $c_1+\cdots+c_s = 0$. Suppose that $s \geq kN(\log K+\log\log K+27)$. Then any set $A \subset [X]^2$ containing only diagonal solutions $(\mathbf{x}_1, \ldots, \mathbf{x}_s) \in A^s$ to the system of equations*

$$c_1\Phi^{u,v}(\mathbf{x}_1) + \cdots + c_s\Phi^{u,v}(\mathbf{x}_s) = 0 \qquad (u + v \geq 0), \qquad (5.1)$$

satisfies the bound

$$|A| \ll X^2 (\log\log X)^{-1/(s-1)}. \qquad (5.2)$$

Here the implicit constant depends only on $\mathbf{c}$ and Φ.

Remark 5.2. We will prove Theorem 5.1 under the assumption that Φ is non-degenerate. The degenerate case follows from the same argument, but the superior bounds available in the standard Vinogradov mean value theorem ensure that, in this case, the lower bound on the number of variables required can be decreased.

In order to prove Theorems 5.1 it is useful to work with translates of sets of the form $[X]^2$. We define a *half-open square* to be a subset of $\mathbb{R}^2$ of the form

$$Q = \mathbf{x} + (0, X]^2,$$

and call X the *side-length* of Q. Let us write $[Q]$ to denote the set of integer points in Q, namely $[Q] = Q \cap \mathbb{Z}^2$.

We reduce the proof of Theorem 5.1 to the following density increment result.

Lemma 5.3. *Given the assumptions in Theorem 5.1, there exist absolute constants $\tau = \tau(k)$, $C = C(\mathbf{c}, \Phi)$ and $c = c(\mathbf{c}, \Phi) > 0$ such that for any $\delta > 0$ and any real $X \geq \exp(C/\delta)$, if $Q \subset \mathbb{R}^2$ is a half-open square with side-length X and $A \subset [Q]$ satisfies $|A| = \delta|[Q]|$ and*

$$R_{\mathbf{c},\Phi}(A) \leq c\delta^s X^{2s-K}, \tag{5.3}$$

then there exists a half-open square Q_1 with side-length at least $2^{-k}X^\tau$, along with $q \in \mathbb{N}$ and $\mathbf{r} \in \mathbb{Z}^2$, such that

$$|A \cap (\mathbf{r} + q \cdot [Q_1])| \geq (\delta + c\delta^s)|[Q_1]|. \tag{5.4}$$

Proof that Lemma 5.3 implies Theorem 5.1. Let us suppose that $A \subset [X]^2$ contains only diagonal solutions to (5.1) and let τ, C and c be as in Lemma 5.3. We aim to construct a sequence of quadruples $(Q_i, A_i, X_i, \delta_i)$ satisfying all of the following conditions.

(i) Q_i is a half-open square of side-length X_i.
(ii) $A_i \subset [Q_i]$ with $A_i = \delta_i|[Q_i]|$.
(iii) A_i contains only diagonal solutions to (5.1).
(iv) $X_{i+1} \geq 2^{-k}X_i^\tau$.
(v) $\delta_{i+1} \geq \delta_i + c\delta_i^s$.

Taking $Q_0 = Q$, $A_0 = A$, $X_0 = X$ and $\delta_0 = |A_0|/|[Q_0]|$, we have our initial quadruple. Let us suppose we have constructed $(Q_j, A_j, X_j, \delta_j)$ for all $1 \leq j \leq i$. In order to apply Lemma 5.3, we must estimate $R_{\mathbf{c},\Phi}(A_i)$. First notice that A_i^s contains exactly $|A_i|$ solutions to (5.1) with $\mathbf{x}_1 = \cdots = \mathbf{x}_s$. Any other solution counted by $R_{\mathbf{c},\Phi}(A_i)$ must have all $\mathbf{x}_j$ contained on some affine line L, where $|L \cap A_i| \geq 2$. Any 2-set $\{\mathbf{x}, \mathbf{y}\} \subset A_i$ is contained in exactly one affine line L. Letting $\mathcal{L}$ denote the set of affine lines which intersect A_i in at least two places, we therefore have

$$\begin{aligned} R_{\mathbf{c},\Phi}(A_i) &\leq |A_i| + \sum_{L \in \mathcal{L}} |L \cap [Q_i]|^s \\ &\leq |A_i| + \binom{|A_i|}{2} \max_{L \in \mathcal{L}} |L \cap [Q_i]|^s \\ &\ll X_i^4 \max_{L \in \mathcal{L}} |L \cap [Q_i]|^s. \end{aligned}$$

The set of integer points in $L \cap Q_i$ projects injectively onto either the x or y axis, with image equal to a set of integer points contained in a subinterval of length X_i. Hence $|L \cap Q_i| \leq X_i + 1$. Thus

$$R_{\mathbf{c},\Phi}(A_i) \ll X_i^{s+4}.$$

Our assumption on the size of s certainly ensures that $2s - K > s + 4$, hence taking C sufficiently large in the assumption

$$X_i \geq \exp(C/\delta_i), \tag{5.5}$$

certainly implies that

$$R_{\mathbf{c},\Phi}(A_i) \leq 2^{s+3}X_i^{s+4} \leq c\delta_i^s X_i^{2s-K}.$$

Assuming (5.5), we can therefore employ Lemma 5.3 to obtain a half-open square Q_{i+1} of side-length $X_{i+1} \geq 2^{-k}X_i^\tau$, together with q and $\mathbf{r}$ such that

$$|A_i \cap (\mathbf{r} + q \cdot [Q_{i+1}])| \geq (\delta_i + c\delta_i^s)|[Q_{i+1}]|.$$

Let us set $A_{i+1} = \{\mathbf{x} \in [Q_{i+1}] : \mathbf{r} + q\mathbf{x} \in A_i\}$ and $\delta_{i+1} = |A_{i+1}|/|[Q_{i+1}]|$. The fact that $c_1 + \cdots + c_s = 0$ means the system (5.1) is translation-dilation invariant. Using this, it follows that if $(\mathbf{x}_1, \dots, \mathbf{x}_s) \in A_{i+1}^s$ is a solution to (5.1), then the tuple $(\mathbf{r}+q\mathbf{x}_1, \dots, \mathbf{r}+q\mathbf{x}_s)$ is a solution to (5.1) in A_i^s. Since A_i has only diagonal solutions to (5.1), it follows that $(\mathbf{x}_1, \dots, \mathbf{x}_s)$ is itself diagonal. Assuming (5.5), we have therefore obtained another quadruple $(Q_{i+1}, A_{i+1}, X_{i+1}, \delta_{i+1})$ satisfying conditions (i) to (v).

As long as (5.5) holds, we can iterate this construction. After $\lceil c^{-1}\delta^{1-s}\rceil$ such iterations we have a density δ_i of size at least 2δ. After a further $\lceil c^{-1}(2\delta)^{1-s}\rceil$ such iterations, we have a density of at least 4δ. Thus, setting $L = \lfloor \log_2(\delta^{-1})\rfloor$, we see that after a total of

$$I = \sum_{l=0}^{L} \left\lceil c^{-1}(2^l\delta)^{1-s}\right\rceil$$

iterations, we have a density of $2^{L+1}\delta > 1$ (a contradiction). Hence (5.5) cannot hold for all $0 \le i \le I$. Thus for some $i \in \{0, 1, \dots, I\}$ we have

$$\begin{aligned}\exp(C/\delta) \ge \exp(C/\delta_i) \ge X_i& \\ &\ge X_{i-1}^{\tau}2^{-k} \ge X_{i-2}^{\tau^2}2^{-k(1+\tau)} \ge \cdots \ge X_0^{\tau^i}2^{-k/(1-\tau)} \\ &\ge X^{\tau^i}2^{-k/(1-\tau)}.\end{aligned} \tag{5.6}$$

Taking logarithms in (5.6), we therefore have

$$C/\delta \ge \tau^i \log X - \tfrac{k}{1-\tau}.$$

Notice that $i \le I \le 2c^{-1}\delta^{1-s}$. So on taking logarithms again we have

$$\log(C\delta^{-1} + k(1-\tau)^{-1}) + 2c^{-1}\delta^{1-s}\log(1/\tau) \ge \log\log X. \tag{5.7}$$

Crude estimation shows that the left hand side of (5.7) is $O_{k,\Phi}(\delta^{1-s})$, as required. □

We begin the proof of Lemma 5.3 with the following general result on partitioning phase polynomials into approximate level sets.

Lemma 5.4. *Let $P(x_1, x_2)$ denote a real polynomial of degree k. Set*

$$\tau_k^{-1} = 24^k(k!)^2 2^{k(k+1)/2}.$$

There exists a positive constant $C = C(k)$, such that for any half-open square Q of side-length X, we can find half-open squares $Q_1, \dots, Q_n$ each with side-length at least $2^{-k}X^{\tau_k}$, along with $q_i \in \mathbb{N}$ and $\mathbf{r}_i \in \mathbb{Z}^2$, such that the sets $\mathbf{r}_i + q_i \cdot [Q_i]$ partition $[Q]$, and furthermore for any $\mathbf{x}, \mathbf{y} \in \mathbf{r}_i + q_i \cdot [Q_i]$ we have

$$\|P(\mathbf{x}) - P(\mathbf{y})\| \le CX^{-\tau_k}. \tag{5.8}$$

Proof. By Taylor's formula

$$\begin{aligned}P(\mathbf{r} + q\mathbf{x}) &= \sum_{u,v\ge 0} \frac{r_1^u r_2^v}{u!v!} P^{u,v}(q\mathbf{x}) \\ &= q^k F(\mathbf{x}) + G(\mathbf{x}; \mathbf{r}, q),\end{aligned} \tag{5.9}$$

where $F(\mathbf{x})$ is a homogeneous real polynomial of degree k, and $G(\mathbf{x}) = G(\mathbf{x}; \mathbf{r}, q)$ is a polynomial of degree strictly less than k. Set

$$F(\mathbf{x}) = \sum_{0\le l\le k} \alpha_l x_1^l x_2^{k-l}, \tag{5.10}$$

and let $\sigma_k^{-1} = 6k2^k$. It follows from Baker [2, Theorem 8.1] that there exists $C_1 = C_1(k)$ such that for any $Y \geq 1$ there is some some $1 \leq q \leq Y$ satisfying

$$\left\|q^k \alpha_l\right\| \leq C_1 Y^{-\sigma_k} \quad (0 \leq l \leq k). \tag{5.11}$$

Let $Y = X^{1/2}$ in (5.11), where X is the side-length of Q. Partitioning $[Q]$ into congruence classes modulo q, we have

$$[Q] = \bigcup_{\mathbf{r} \in [q]^2} \mathbf{r} + q \cdot [Q(\mathbf{r})],$$

where $Q(\mathbf{r})$ is a half-open square of side-length X/q. Let us set

$$t = \left\lceil q^{-1} X^{1-\frac{\sigma_k}{4k}} \right\rceil.$$

Then we can partition each $Q(\mathbf{r})$ into t^2 half-open squares $Q(\mathbf{r}, \mathbf{t})$ ($\mathbf{t} \in [t]^2$), each of side-length $X/(qt)$. For fixed $\mathbf{r}$ and $\mathbf{t}$ let us pick $\mathbf{a}(\mathbf{r}, \mathbf{t}) \in [Q(\mathbf{r}, \mathbf{t})]$. Then we have $[Q(\mathbf{r}, \mathbf{t})] = \mathbf{a}(\mathbf{r}, \mathbf{t}) + [Q'(\mathbf{r}, \mathbf{t})]$, where $Q'(\mathbf{r}, \mathbf{t})$ is a half-open square of side-length $X/(qt)$ satisfying

$$Q'(\mathbf{r}, \mathbf{t}) \subset [-X/(qt), X/(qt)]^2. \tag{5.12}$$

It follows that there exist pairs $\mathbf{b}(\mathbf{r}, \mathbf{t}) \in \mathbb{Z}^2$ ($\mathbf{r} \in [q]^2$, $\mathbf{t} \in [\mathbf{T}]$) such that the set $[Q]$ is equal to the disjoint union

$$\bigcup_{\mathbf{r} \in [q]^2} \bigcup_{\mathbf{t} \in [\mathbf{T}]} \left(\mathbf{b}(\mathbf{r}, \mathbf{t}) + q \cdot [Q'(\mathbf{r}, \mathbf{t})]\right).$$

Clearly $X/(qt) \leq X^{\sigma_k/(4k)}$. Since $q \leq X^{1/2} \leq X^{1-\sigma_k/(4k)}$, we also have

$$X/(qt) \geq \frac{X}{X^{1-\sigma_k/(2k)} + q} \geq \tfrac{1}{2} X^{\sigma_k/(4k)}.$$

Hence the side-length of each $Q'(\mathbf{r}, \mathbf{t})$ is between $\frac{1}{2} X^{\sigma_k/(4k)}$ and $X^{\sigma_k/(4k)}$. It follows that for any $\mathbf{x}, \mathbf{y} \in [Q'(\mathbf{r}, \mathbf{t})]$ we have

$$\left\|q^k (F(\mathbf{x}) - F(\mathbf{y}))\right\| \ll_k X^{-\sigma_k/4}. \tag{5.13}$$

Write $G_{\mathbf{r},\mathbf{t}}(\mathbf{x})$ for the polynomial $G(\mathbf{x}; \mathbf{b}(\mathbf{r}, \mathbf{t}), q)$. By induction, there exists a partition of $[Q'(\mathbf{r}, \mathbf{t})]$ into sets of the form $\mathbf{s}_i + q_i \cdot [Q_i'(\mathbf{r}, \mathbf{t})]$ ($1 \leq i \leq m = m(\mathbf{r}, \mathbf{t})$), where $Q_i'(\mathbf{r}, \mathbf{t})$ is a half-open square of side-length at least

$$2^{1-k-\tau_{k-1}} X^{\sigma_k \tau_{k-1}/(4k)},$$

and such that for any $\mathbf{x}, \mathbf{y} \in [Q_i'(\mathbf{r}, \mathbf{t})]$ we have

$$\|G_{\mathbf{r},\mathbf{t}}(\mathbf{s}_i + q_i \mathbf{x}) - G_{\mathbf{r},\mathbf{t}}(\mathbf{s}_i + q_i \mathbf{y})\| \ll_k X^{-\sigma_k \tau_{k-1}/(4k)}. \tag{5.14}$$

Let us write $q_i'(\mathbf{r}, \mathbf{t})$ for $q q_i(\mathbf{r}, \mathbf{t})$ and $\mathbf{b}_i'(\mathbf{r}, \mathbf{t})$ for $\mathbf{b}(\mathbf{r}, \mathbf{t}) + q \mathbf{s}_i(\mathbf{r}, \mathbf{t})$. Then $[Q]$ is partitioned by the sets

$$\mathbf{b}_i'(\mathbf{r}, \mathbf{t}) + q_i'(\mathbf{r}, \mathbf{t}) \cdot [Q_i'(\mathbf{r}, \mathbf{t})] \quad (\mathbf{r} \in [q]^2,\ \mathbf{t} \in [\mathbf{T}],\ 1 \leq i \leq m(\mathbf{r}, \mathbf{t})).$$

Combining (5.9), (5.13) and (5.14), we see that for each $\mathbf{x}, \mathbf{y} \in [Q_i'(\mathbf{r}, \mathbf{t})]$ we have, on writing $\mathbf{b}' = \mathbf{b}_i'(\mathbf{r}, \mathbf{t})$ and $q' = q_i'(\mathbf{r}, \mathbf{t})$, that

$$\left\|P(\mathbf{b}' + q'\mathbf{x}) - P(\mathbf{b}' + q'\mathbf{y})\right\| \ll X^{-\sigma_k \tau_{k-1}/(4k)}.$$

A simple calculation reveals that $\sigma_k \tau_{k-1}/(4k) = \tau_k$, as required. □

Proof of Lemma 5.3. Let us define

$$f_A(\boldsymbol{\alpha}) = \sum_{\mathbf{x}} 1_A(\mathbf{x}) e(\boldsymbol{\alpha} \cdot \mathbf{F}(\mathbf{x})),$$

together with $f(\boldsymbol{\alpha}) = f_{[Q]}(\boldsymbol{\alpha})$ and $g_A(\boldsymbol{\alpha}) = f_A(\boldsymbol{\alpha}) - \delta f(\boldsymbol{\alpha})$. Using translation invariance of (5.1), we have $R_{\mathbf{c},\Phi}([Q]) \sim R_{\mathbf{c},\Phi}(X)$. Let c_1 equal the quantity $\sigma_\infty(\mathbf{c}) \prod_p \sigma_p(\mathbf{c})$ from Theorem 4.3. Provided we take C in Lemma 5.3 sufficiently large, so that $X \geq \exp(C/\delta^{-1}) \gg_{\mathbf{c},\Phi} 1$, we can use Theorem 4.3 to ensure that

$$R_{\mathbf{c},\Phi}([Q]) \geq \tfrac{1}{2} c_1 X^{2s-K}. \tag{5.15}$$

The assumption that $\mathbf{c}$ is a non-singular choice for Φ implies that c_1 is positive. Let us take c in Lemma 5.3 sufficiently small, say $c \leq \frac{1}{4} c_1$. Combining (5.3), (5.15), orthogonality and Hölder's inequality, we have

$$\begin{aligned} \tfrac{1}{4} c_1 \delta^s X^{2s-K} &\leq |R_{\mathbf{c},\Phi}(A) - \delta^s R_{\mathbf{c},\Phi}([Q])| \\ &\leq \oint |f_A(c_1\boldsymbol{\alpha}) \cdots f_A(c_s\boldsymbol{\alpha}) - \delta^s f(c_1\boldsymbol{\alpha}) \cdots f(c_s\boldsymbol{\alpha})| \mathrm{d}\boldsymbol{\alpha} \\ &\leq \sup_{\boldsymbol{\alpha}} |g_A(\boldsymbol{\alpha})| X^{2\epsilon} \oint \big(|f_A(\boldsymbol{\alpha})|^{2t} + |f(\boldsymbol{\alpha})|^{2t}\big) \mathrm{d}\boldsymbol{\alpha}, \end{aligned}$$

where $s = 1 + \epsilon + 2t$, for some $\epsilon \in \{0, 1\}$. Since $2t \geq kN(\log K + \log\log K + 26)$, we can use Theorem 4.3 (together with the underlying Diophantine equation), to conclude that

$$\begin{aligned} \oint |f_A(\boldsymbol{\alpha})|^{2t} \mathrm{d}\boldsymbol{\alpha} \leq \oint |f(\boldsymbol{\alpha})|^{2t} \mathrm{d}\boldsymbol{\alpha} \\ \ll_{\mathbf{c},\Phi} X^{4t-K}. \end{aligned}$$

Setting $b_A(\mathbf{x}) = 1_A(\mathbf{x}) - \delta 1_{[Q]}(\mathbf{x})$, we see that there exists $\boldsymbol{\alpha} \in \mathbb{T}^N$ such that

$$\Big| \sum_{\mathbf{x} \in [Q]^2} b_A(\mathbf{x}) e(\boldsymbol{\alpha} \cdot \mathbf{F}(\mathbf{x})) \Big| = |g_A(\boldsymbol{\alpha})| \gg_{\mathbf{c},\Phi} \delta^s X^2. \tag{5.16}$$

Let $\tau = \tau_k$ and $C_1 = C_1(k)$ be as in Lemma 5.4, and consider the polynomial $P(\mathbf{x}) = \boldsymbol{\alpha} \cdot \mathbf{F}(\mathbf{x})$. Then there exist half-open squares $Q_1, \dots, Q_n$ each of side-length at least $2^{-k} X^\tau$, along with $\mathbf{r}_i$ and q_i $(1 \leq i \leq n)$ such that the sets $\mathbf{r}_i + q_i \cdot [Q_i]$ partition $[Q]$, and for any $\mathbf{x}, \mathbf{y} \in \mathbf{r}_i + q_i \cdot [Q_i]$ we have $\|P(\mathbf{x}) - P(\mathbf{y})\| \leq C_1 X^{-\tau}$. Notice that this implies that $|e(P(\mathbf{x}) - e(P(\mathbf{y}))| \ll X^{-\tau}$. Thus

$$\begin{aligned} \Big| \sum_{\mathbf{x} \in [X]^2} b_A(\mathbf{x}) e(\boldsymbol{\alpha} \cdot \mathbf{F}(\mathbf{x})) \Big| &\leq \sum_{i=1}^n \Big| \sum_{\mathbf{x} \in \mathbf{r}_i + q_i \cdot [Q_i]} b_A(\mathbf{x}) e(P(\mathbf{x})) \Big| \\ &= \sum_{i=1}^n \Big| \sum_{\mathbf{x} \in \mathbf{r}_i + q_i \cdot [Q_i]} b_A(\mathbf{x}) \Big| + O\big(X^{2-\tau}\big). \end{aligned}$$

We can take C in Lemma 5.3 sufficiently large to ensure that the lower bound $X \geq \exp(C/\delta)$ implies that the $O(X^{2-\tau})$ term above is at most half the size of the right hand side of (5.16). We thereby obtain that

$$\sum_{i=1}^n \Big| \sum_{\mathbf{x} \in \mathbf{r}_i + q_i \cdot [Q_i]} b_A(\mathbf{x}) \Big| \gg_{\mathbf{c},\Phi} \delta^s X^2. \tag{5.17}$$

Let $\mathcal{I}$ denote the set of $i \in [n]$ for which $\sum_{\mathbf{x}\in\mathbf{r}_i+q_i\cdot[Q_i]} b_A(\mathbf{x}) \geq 0$. Since b_A has average zero, we can add $\sum_{\mathbf{x}} b_A(\mathbf{x})$ to the left side of (5.17), to obtain

$$\sum_{i\in\mathcal{I}}\Big(\sum_{\mathbf{x}\in\mathbf{r}_i+q_i\cdot[Q_i]} b_A(\mathbf{x})\Big) \gg_{\mathbf{c},\Phi} \delta^s X^2. \tag{5.18}$$

The density increment (5.4) now follows from the pigeon-hole principle, provided we take $c = c(\mathbf{c}, \Phi)$ sufficiently small. □

Our originally advertised theorem, Theorem 1.3, now almost follows. It remains to show that $kN(\log K + \log\log K + 27) \leq \frac{3}{4}k^3 \log k(1 + o(1))$. We have the trivial bound $N \leq k^2$ and $K \leq kN \leq k^3$. The bound $N \leq \frac{k^2}{4}(1 + o(1))$ takes a little more calculation, but follows from the fact that the number of linearly independent derivatives $\Phi^{u,v}$ with $u + v = k - d$ is at most $\max\{k + 1 - d, d + 1\}$.

Acknowledgements. The author would like to express his gratitude to Professor Wooley for his unending encouragement, patience and generosity with ideas, and Professor Parsell for his insights into §3.

References

1. G. I. Arhipov, A. A. Karacuba, and V. N. Čubarikov, *Multiple trigonometric sums*, Proc. Steklov Inst. Math. (1982), no. 2, viii+126, A translation of Trudy Mat. Inst. Steklov. **1**51 (1980).
2. R. C. Baker, *Diophantine inequalities*, London Mathematical Society Monographs. New Series, vol. 1, The Clarendon Press Oxford University Press, New York, 1986, Oxford Science Publications.
3. J. Bourgain, *On triples in arithmetic progression*, Geom. Funct. Anal. **9** (1999), no. 5, 968–984.
4. ———, *Roth's theorem on progressions revisited*, J. Anal. Math. **104** (2008), 155–192.
5. H. Furstenberg and Y. Katznelson, *An ergodic Szemerédi theorem for commuting transformations*, J. Analyse Math. **34** (1978), 275–291 (1979).
6. W. T. Gowers, *A new proof of Szemerédi's theorem*, Geom. Funct. Anal. **11** (2001), no. 3, 465–588.
7. B. J. Green and T. Tao, *New bounds for Szemerédi's theorem. II. A new bound for $r_4(N)$*, Analytic number theory, Cambridge Univ. Press, Cambridge, 2009, pp. 180–204.
8. H. Iwaniec and E. Kowalski, *Analytic number theory*, American Mathematical Society Colloquium Publications, vol. 53, American Mathematical Society, 2004.
9. A. A. Karatsuba, *The mean value of the modulus of a trigonometric sum*, Izv. Akad. Nauk SSSR **37** (1973), 1203–1227.
10. U. V. Linnik, *On Weyl's sums*, Rec. Math. [Mat. Sbornik] N.S. **12(54)** (1943), 28–39.
11. H. L. Montgomery, *Ten lectures on the interface between analytic number theory and harmonic analysis*, CBMS Regional Conference Series in Mathematics, vol. 84, Published for the Conference Board of the Mathematical Sciences, Washington, DC, 1994.
12. S. T. Parsell, *Pairs of additive equations of small degree*, Acta Arith. **104** (2002), no. 4, 345–402.
13. ———, *A generalization of Vinogradov's mean value theorem*, Proc. London Math. Soc. (3) **91** (2005), no. 1, 1–32.
14. K. F. Roth, *On certain sets of integers*, J. London Math. Soc. **28** (1953), 104–109.
15. ———, *On certain sets of integers. II*, J. London Math. Soc. **29** (1954), 20–26.
16. W. M. Schmidt, *Simultaneous rational zeros of quadratic forms*, Seminar on Number Theory, Paris 1980-81 (Paris, 1980/1981), Progr. Math., vol. 22, Birkhäuser Boston, Mass., 1982, pp. 281–307.
17. I. D. Shkredov, *On a generalization of Szemerédi's theorem*, Dokl. Akad. Nauk **405** (2005), no. 3, 315–319.
18. M. L. Smith, *On solution-free sets for simultaneous additive equations*, ProQuest LLC, Ann Arbor, MI, 2007, Thesis (Ph.D.)–University of Michigan.
19. ———, *On solution-free sets for simultaneous quadratic and linear equations*, J. Lond. Math. Soc. (2) **79** (2009), no. 2, 273–293.

20. S. B. Stechkin, *On mean values of the modulus of a trigonometric sum*, Trudy Mat. Inst. Steklov **134** (1975), 283–309.
21. R. C. Vaughan, *The Hardy-Littlewood method*, second ed., Cambridge Tracts in Mathematics, vol. 125, Cambridge University Press, Cambridge, 1997.
22. T. D. Wooley, *A note on simultaneous congruences*, J. Number Theory **58** (1996), no. 2, 288–297.
23. ———, *On Weyl's inequality, Hua's lemma, and exponential sums over binary forms*, Duke Math. J. **100** (1999), no. 3, 373–423.

School of Mathematics, University of Bristol, University Walk, Clifton, Bristol BS8 1TW, United Kingdom

E-mail address: sean.prendiville@bristol.ac.uk